THE INFLUENCE OF CRITICAL ASSET MANAGEMENT FACETS ON IMPROVING RELIABILITY IN POWER SYSTEMS

A Thesis
Presented to
The Academic Faculty

by

Joshua Perkel

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
December 2008

THE INFLUENCE OF CRITICAL ASSET MANAGEMENT FACETS ON IMPROVING RELIABILITY IN POWER SYSTEMS

Approved by:

Dr. Miroslav Begovic, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Thomas Michaels
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Ronald Harley
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Nigel Hampton
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Brani Vidakovic
Department of Biomedical Engineering
Georgia Institute of Technology
School of Medicine
Emory University

Date Approved: October 21, 2008

To my loving family and reasons for being, Yamille and Elizabeth.

ACKNOWLEDGEMENTS

This thesis represents the conclusion of a journey that began nearly six years ago when I first arrived at Georgia Tech. It has been a long adventure with many moments of doubt and frustration interspersed amongst the moments of celebration for a long fought success. Many individuals took it upon themselves to aid me in my journey, to these individuals I will be eternally grateful.

I would like to begin with the gentleman that started me on my adventure, Dr. Miroslav Begovic. Dr. Begovic and I met roughly one week after my arrival in Atlanta. I was assigned to him as a teaching assistant. During that first semester it became clear that we worked well together and so tried the following semester to work together again. But that was not to be. At that time I was not planning to pursue a PhD, but after a few months I became intrigued by the idea of research and so when it was time to select my advisor I knew exactly whose door to knock on. In the years since then Dr. Begovic has become a mentor, colleague, and friend, without whom I could not have come this far. Thank you for all your support and patience.

None of this would have been possible without the creative, high-flying, thinking of Dr. Nigel Hampton of the National Electrical Energy Testing, Research, and Applications Center (NEETRAC). Dr. Hampton is one of the most original thinkers I have ever met. His enthusiasm and energy are powerful motivators that have inspired me in the moments when I was facing the proverbial "brick wall." His insight and experience have been vital contributions to the preparation of this thesis. Moreover, his support and understanding have made this entire journey possible. Thank you for everything Nigel.

Yet another individual whose support has been vital for me is Mr. Rick Hartlein of NEETRAC. He and I have worked on projects together for the last several years including the Cable Diagnostic Focused Initiative (CDFI). This project introduced me to

the world of cable diagnostics and it formed the basis for much of the work presented in this thesis. Rick is the sort of leader that enriches the work of those he leads and I am truly grateful for the opportunity to be one of those lucky few. Thank you Chief.

I would also like to extend my appreciation to the gentlemen and distinguished faculty that graciously agreed to serve on my PhD committee. Dr. Ronald Harley, Dr. Thomas Michaels, and Dr. Brani Vidakovic, thank you for your support and advice.

Moreover, I would like to acknowledge the support of NEETRAC and the School of Electrical and Computer Engineering (ECE) for providing the financial support for my MS and PhD graduate studies. Individuals such as Mr. Tim Andrews, Mr. Thomas Parker, Mr. Jorge Altamirano, and Mrs. Gail Reeves from NEETRAC, each deserve my thanks and appreciation for their help and encouragement.

None of this would have been possible without the support of my friends and family. While the pieces of this thesis were taking shape so was my future family. My "brother," Jean Carlos Hernandez-Mejía, has stood by me through each of the difficult moments both in work and in life. He has been one of the inspirations behind my PhD. So have my Yamille and, recently, our Elizabeth. Together these ladies have inspired and motivated me to accomplish what I thought would take forever. They are the "why" and "how" for everything I do. Without them, I would never have come as far as I have and it is for this reason that I dedicate this thesis to them.

Once more, to all those who have been part of this process, please accept my humblest and sincerest thanks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The objective of the proposed research is to develop statistical algorithms for controlling failure trends through targeted maintenance of at-risk components. The at-risk components are identified via chronological history and diagnostic data, if available. Utility systems include many thousands (possibly millions) of components with many of them having already exceeded their design lives. Unfortunately, neither the budget nor manufacturing resources exist to allow for the immediate replacement of all these components. On the other hand, the utility cannot tolerate a decrease in reliability or the associated increased costs. To combat this problem, an overall maintenance model has been developed that utilizes all the available historical information (failure rates and population sizes) and diagnostic tools (real-time conditions of each component) to generate a maintenance plan. This plan must be capable of delivering the needed reliability improvements while remaining economical. It consists of three facets each of which addresses one of the critical asset management issues:

- Failure Prediction Facet – Statistical algorithm for predicting future failure trends and estimating required numbers of corrective actions to alter these failure trends to desirable levels. Provides planning guidance and expected future performance of the system.
- Diagnostic Facet – Development of diagnostic data and techniques for assessing the accuracy and validity of that data. Provides the true effectiveness of the different diagnostic tools that are available.
- Economics Facet – Stochastic model of economic benefits that may be obtained from diagnostic directed maintenance programs. Provides the cost model that may be used for budgeting purposes.

These facets function together to generate a diagnostic directed maintenance plan whose goal is to provide the best available guidance for maximizing the gains in reliability for the budgetary limits utility engineers must operate within.

CHAPTER 1: INTRODUCTION AND BACKGROUND

1.1 INTRODUCTION

The electric power system infrastructure in the United States received large influxes of capital investment in the 1950-1970s for the purpose of expanding the system to accommodate the growing demand by consumers and industry for electricity. During this period of time, tremendous numbers of components were installed with expected service lives of around 20 years. Unfortunately, today, many utility systems still rely on these same components to provide reliable service even though they are now well beyond their original design lives. To make matters worse, this infrastructure is now being called upon to deliver more electric power than ever before. The potential danger in operating an aging system closer and closer to its operational limits increases to the point where wide area blackouts, such as the one that occurred in August 2003 in the northeastern United States, could become commonplace.

The goal of each and every utility is to avoid such outages. However, the utility strives to do so while earning the highest possible profits. During the 1990s, this second goal was given higher priority as maintenance programs were, in general, reduced significantly during this period. The failure rates for many component types were still relatively low during this period and so there was no urgent reason to provide significant investment in the infrastructure. The de-regulation of the utility industry further exacerbated this problem as the restructuring of vertically integrated utilities into separate entities for generation, transmission, and distribution, made planning too uncertain to act on. In general, this restructuring made utilities more cautious about their spending, especially their spending on reliability [1].

The term *reliability* is used to describe the likelihood that a component or system will adequately perform its intended function for a specified period of time. Maintaining reliable operation of an electric grid means the utility must maintain the supply of electricity to each customer. When the reliability of the power system declines, customers experience outages as a result. These outages lead to a "reactive" response from the utilities rather than a well planned approach to managing their aged components.

The conservative spending approach to reliability has now caught up with many utilities. The failure rates [failures/component/year] for many components are increasing at an exponential rate and managers are facing a dilemma of where to allocate the (often very limited) resources for the best possible use [2], [3]. Eventually, all components within the power system will need to be replaced. However, they cannot all be replaced at the same time since neither the funds nor the manufacturing facilities exist for such a total replacement program. On the other hand, the sheer number of components makes action now necessary. Given these obstacles there is still a bright side in which a solution may be found. The fact is that components do not age uniformly throughout a utility system. Many factors influence the lifetime of a component, only one of which is age. In the case of underground cables, factors such as installation method, soil condition, loading, average temperature, maintenance history, and others cause the lifetime for cables to be highly variable even for the same age and design of cable [4], [5]. This fact provides the most valuable commodity for utilities: time to act.

1.2 PROBLEM STATEMENT

The primary problem faced by utilities and the topic of this research is to define a strategy for how to best utilize the time and available budget to achieve improved reliability. The fact that aging and the component's time to failure are unique for individual components that were otherwise identical when they were installed represents the means by which the reliability of the system may be managed. If the utility had the

ability to identify those components that had nearly aged to the point of failure then they could be removed from service before causing a failure in service. Such an approach would maximize the useful life of the component while at the same time maintaining a high level of reliability. This type of strategy falls under the general heading of asset management [6]-[8]. The concept of asset management was originally introduced in the financial world as a means of managing financial assets, but it has since been applied in other areas [9]. Kostic defines asset management as "... the process of guiding the acquisition, use and disposal of assets to make the most of their future economic benefit and manage the related risks and costs over their entire life [9]." This is but one approach to defining asset management. Yet another is the separation proposed by Brown and Spare, shown graphically in Figure 1.

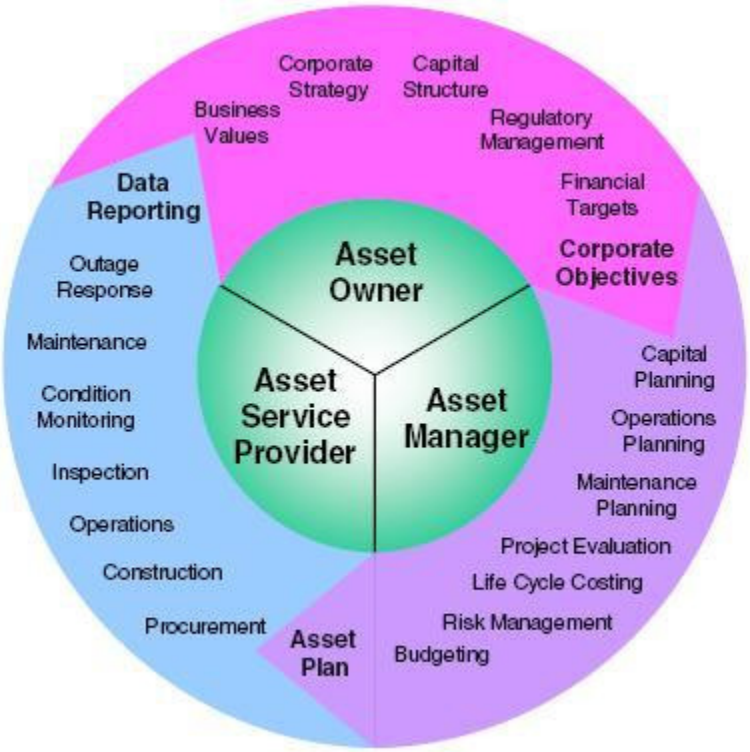


Figure 1: Asset management framework proposed by Brown and Spare [10].

Figure 1 shows that there are three stakeholders in the asset management framework: (1) asset owner, (2) asset service provider, and (3) asset manager. These stake holders may, in fact, all be part of the same organization. Note that the asset owner supplies the corporate objectives that must be considered in the asset plan produced by the asset manager. Meanwhile, the asset service provider supplies the system data and performs the maintenance on the system.

As mentioned, for the purpose of managing aging equipment, the goal is to extract the maximum useful life from each piece of equipment while at the same time avoiding service failures. But asset managers have other concerns as well. These include achieving financial objectives, meeting corporate goals, and creating multi-year plans that will consistently meet all of these objectives [10]. Within this research, the focus is on achieving the technical goals at the lowest expense. Doing so requires a thorough analysis of the available resources, constraints, and objectives.

1.3 PROBLEM ORIGIN AND HISTORY

In the last decade, the electric power industry has become very interested in the development of asset management strategies for handling its aging equipment problem. Implementing an asset management strategy in an actual power system requires a number of tools including the following:

- Failure Prediction – Means of modeling and predicting the reliability of different component populations.
- Maintenance – Method of acting on aged components to restore their operation to "like new" condition, if possible.
- Diagnostics – Techniques for assessing the state or condition of a component at any particular point in its lifetime.

- Economics – Model that predicts the savings and reliability improvement that can be achieved with a chosen maintenance plan.

These critical asset management components are discussed in the following sections.

1.3.1 Failure Prediction

All methods of failure prediction involve either statistical or heuristic procedures, each of which assumes that a common failure mechanism operates on all the components in the population. In this discussion, the population of components is assumed to be homogeneous in the sense that all components are of the same basic type. Each of the techniques attempts to identify a set of model parameters that can then be used to predict future failures. The statistical techniques include distribution fitting [11]-[13], Crow-AMSAA [14], Laplace trend statistic [15], time series analysis [16], Bayesian technique [17]-[18], and Markov Chains [19]-[20]. On the heuristic side, artificial intelligence techniques, such as the neural network, have been employed [21]. These techniques are each briefly described below:

1.3.1.1 Distribution Fitting

This technique models the failures through a parametric distribution that is fitted to time-to-failure data for a specific component type. Various parametric distributions have been used, including normal, Weibull, exponential, Poisson, and a variety of others [11]-[13]. Predictions can be made based in the parameters of the chosen distribution and the number of components in the population.

1.3.1.2 Crow-AMSAA

The Crow-AMSAA technique, or reliability growth model, was developed from the Weibull distribution. It allows predictions of future service performance to be made

based on the trend of cumulative service failures-versus-time [14]. This technique can be applied in situations where the population data is not readily available.

1.3.1.3 Laplace Trend Statistic

Kim et al. propose using the Laplace test statistic as a means of analyzing incipient fault data [15]. With this technique, the goal is to predict service failures in real time perhaps seconds before they actually occur.

1.3.1.4 Time Series Analysis

Standard time series analysis techniques based on regression such as autoregressive moving average (ARMA) or autoregressive integral moving average (ARIMA) can also be used provided a sufficiently large dataset is available [16]. Like the Crow-AMSAA technique, this technique can be used when population information is not available.

1.3.1.5 Bayesian

The Bayesian technique is utilized in conjunction with a parametric distribution as part of an evolutionary process. The process is evolutionary since new data are incorporated into the existing parametric distribution (prior) to obtain a new posterior distribution that is augmented with the newest data. Examples of this technique in practice include prognostics for aerospace, as described by Engel et al. [17], and prediction of transformer failures by Gulacchenski and Besuner [18].

1.3.1.6 Markov Chains

Markov chains are employed where discrete states or conditions for the component can be defined. This technique can be used to model the progression of a component's degradation process. Each stage of this process would correspond to a unique state. As described by Endrenyi [19]-[20], each of these states has associated with it a set of

transition probabilities to all other states. Once given a starting state it is possible to then follow the chain using the transition probabilities.

1.3.1.7 Neural Network

Chinnam describes an online procedure that defines failure in terms of a specified level of degradation and then estimates the reliability of that component based on that level of degradation [21]. The technique utilizes a self-organizing map to capture the variations in the degradation and a feed-forward neural network to determine the degradation measures [21].

Each of the above techniques may be employed to perform a failure prediction. However, for this research, a statistical technique has been chosen and is described in Part I.

1.3.2 Maintenance

Failure predictions allow one to determine the reliability trends of the system so that plans can be made for disrupting an increasing loss of reliability. Once the time is appropriate for corrective actions to be taken, the utility has several maintenance strategies at its disposal. The term "maintenance" is defined by Endrenyi as "a restoration wherein an un-failed device has, from time to time, its deterioration arrested, reduced, or eliminated [19]." Corrective actions would include both maintenance actions and replacement actions.

There are three basic maintenance policies employed by utilities today:

- Corrective Maintenance
- Preventive Maintenance
- Predictive Maintenance

Each of these forms of maintenance requires a different amount of routine observation and other ongoing work on the part of the utility. This translates into

different levels of financial commitment to the maintenance program and, if things go as planned, different levels of reliability.

The first type, corrective maintenance, is the simplest approach in that the utility simply repairs or replaces a component only once it has failed in service. Given this simple principle, this method is also known as “repair on failure” or “run to failure.” This method requires no attention from the utility other than when a failure has occurred. While this results in low operational cost, it produces the poorest reliability as all components fail in service as a result of the program.

The second type, preventive maintenance, involves the use of scheduled maintenance periods in which the component is partially or completely restored to new condition [20]. These maintenance cycles occur regardless of the condition of the component. An example of such a policy is the regular oil change in an engine. Generally, this policy uses fixed intervals that are determined from an analysis of the available failure data. As the component may not be on the verge of failure, this type of maintenance expends considerable financial resources on maintenance performed too regularly and indiscriminately.

The last maintenance policy, predictive maintenance, utilizes condition monitoring data to trigger maintenance actions only once the component is determined to be in a deteriorated state [1]. This policy has the potential to be superior to the others in terms of both reliability and cost since maintenance is only performed when needed, but is completed before a service failure occurs. Unfortunately, such a condition-based approach requires some means to monitor the condition of the device either in real time or periodically through inspections or diagnostic tests. As a consequence, the monitoring process does consume maintenance funds. The advantage, though, is that tracking the component’s condition over time allows the utility to observe changes in condition, which are considerably better indicators of deterioration than a single “bad” measurement

[1]. An alternative approach is reliability centered maintenance (RCM) in which, in addition to condition monitoring data, the cost and criticality are included as parameters for determining which components will receive maintenance [22].

Each of these methods is employed in utilities today. However, the corrective maintenance policy is still predominant because of the low operational cost and lack of condition monitoring data. Utilities, unfortunately, have been poor record keepers over the years and altering this habit has been a challenging task.

1.3.3 Diagnostics

Diagnostics has become an area of intense interest by both utilities and the US Department of Energy (DoE) as a way to alleviate some of the reliability issues in the power system. In fact, the DoE currently co-sponsors a focused initiative through the National Electric Energy Testing and Research Applications Center (NEETRAC) at Georgia Tech in the area of underground cable diagnostics. This project is known as the Cable Diagnostics Focused Initiative (CDFI). Much of the author's research is based on work performed as part of this project, so the discussion on diagnostics will be driven by those diagnostics that are relevant for underground cables. Other component types utilize different diagnostic techniques but the use and interpretation of the diagnostic data will be similar. The methods and models developed as part of this research are not limited to any class of component or diagnostic technique.

Diagnostic tests are designed to measure component characteristics that can provide information as to the level of deterioration the component has suffered with respect to a specific failure mechanism. In other words, diagnostics do not directly measure deterioration or degradation. They instead look for signs that indicate that deterioration of the type that would be expected for a specific failure mechanism is present. Such tests can be local or global in nature in the sense that local tests can locate specific points of deterioration within the component, while global techniques are able to provide the

general condition of the component [4]. An example of a local technique for underground cables is the partial discharge test, while a global condition assessment can be made using the dissipation factor test [5], [25], [26]. Depending on the chosen maintenance actions, one test type may be more appropriate than another.

If diagnostic tests are to be used in the asset management process then their associated accuracies must be known. It is well known that diagnostic tests are less than 100% accurate but exactly how much less than 100% still allows them provide useful information is another question. The diagnostic should only be used if the risk of failure under test is acceptable to the asset manager. Furthermore, the diagnostic program must be more cost effective than alternative programs. This also depends on the level of accuracy that the diagnostic technique is able to deliver. The literature discusses much about the individual techniques themselves but fails to mention the level of accuracy that these tests deliver.

1.3.4 Economics

The economics of reliability programs must be evaluated by considering both the costs involved in the maintenance program and the associated improvement or reduction in reliability. The cost components can be calculated through the number and extent of the corrective actions. Meanwhile, reliability changes can be observed by the numbers of failures or by the resulting reliability indices. This type of analysis works well for program evaluation. However, often times the goal is to develop a maintenance plan that is subject to a constrained budget. Brown and Marshall [27] have developed a methodology that is able to accomplish this objective. In addition, Painton and Campbell [26] have used the genetic algorithm to produce an optimized maintenance plan.

According to Kostic [9], in reality, the utility must perform two tasks concerning economics:

- Prioritize expenditures in operation and maintenance (O&M).

- Prioritize capital spending for expansion and component replacement.

It must be noted that each of the above tasks depends on the available capital and O&M budgets, which can vary substantially from year to year.

The issue of how to compute the costs for an asset management driven maintenance plan is discussed in detail in Part III.

1.4 SUMMARY OF BACKGROUND MATERIAL

This chapter describes the work done by others in the field of asset management applied to aging power equipment. It is clear that significant work has been done to create failure prediction algorithms. These algorithms provide point estimates of future failures. Also, a number of maintenance policies are currently in use in the industry, with the condition monitoring approach being the most sophisticated. To implement such a maintenance policy, it is necessary to perform routine inspections or diagnostic tests to evaluate if and what level of maintenance is necessary to keep a component operational. Of course, these decisions must be based on methods that are less than perfect in terms of accuracy, but even so they may be the best alternative for achieving economical and reliable operation of the power system.

1.5 RESEARCH PERSPECTIVE

The objective of this research is to develop a statistical algorithm for controlling failure trends through the targeted maintenance of components that are at-risk for failure. Electric utilities own and operate vast networks of components that must function in harmony for electric power to be generated and delivered safely to customers. Unfortunately, each of these devices possesses a finite but unknown lifetime during which it is able to perform the required function. Occasionally, these devices fail while in service. As a result, utilities must devote significant portions of their annual operating budgets to the maintenance and care of their equipment if they are to continue to

successfully conduct business. The term "asset management" is generally used to describe the process of planning, budgeting, procuring, data logging, maintenance scheduling, and follow-up needed to optimally utilize the available resources both in the present and in the future. Restated another way, asset management is about keeping the power network operational while maximizing the return for every dollar spent. It is an optimization problem based on achieving specific high-priority goals using the best (often incomplete) information available. Asset management works at all levels within a utility and is too large a topic to tackle all at once; therefore, it is necessary to focus on a specific part of the utility. The modeling and discussion that are contained within this thesis focus on the challenge of managing the aged equipment within the utility.

This problem may be approached from two different perspectives:

- Given a finite and known amount of resources (budget), determine the most effective means of distributing those resources so as to maximize the improvement in reliability of the aged component population.
- Given a specified level of reliability, determine the minimum budget needed to guarantee that the performance of the aged components meets or exceeds the desired reliability.

While the inputs of these two approaches are different, their ultimate goal is the same: to achieve the best system performance possible using all the available resources. This "all" includes the current and future financial resources as well as the components currently in service and those that will be put into service. The in-service components can have associated with them two basic types of information:

- Historical Data – Characterizes the performance of the components in the past so that inferences may be made as to their future performance.
- Condition Data – Characterizes the current "health" of each specific component using a diagnostic test.

In the past, utilities have focused on one type of data or the other. However, used separately each data type's usefulness is limited and, as will be shown in the following chapters, both can be used to achieve better performance. Figure 2 shows conceptually how an overall decision process may be constructed that would include inputs from an *aging system* to three analysis and processing facets: (1) failure prediction facet, (2) diagnostic facet, and (3) economics facet.

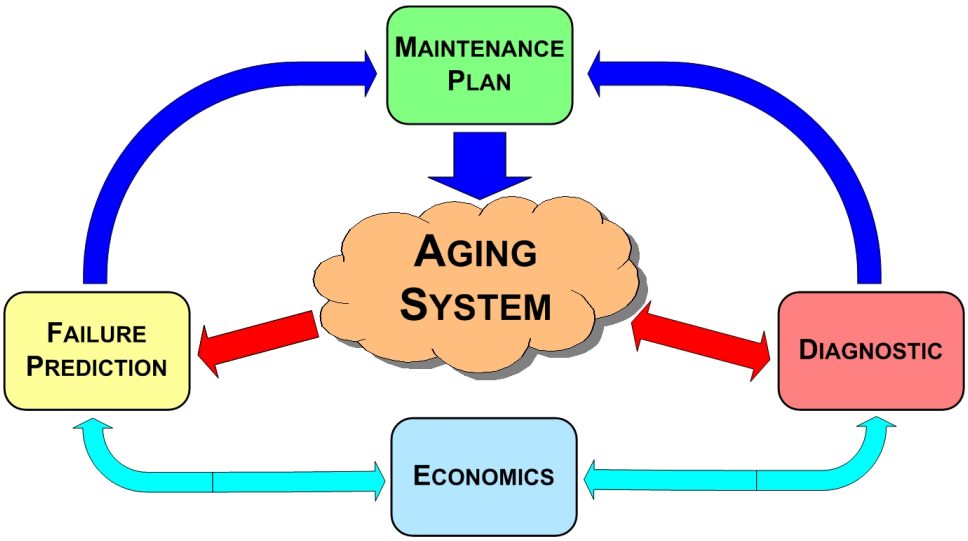


Figure 2: Basic flow chart of the three major components of the asset management process for creating a maintenance plan.

Each facet can be thought of as one stage of filtration that begins with the entire component population. The *failure prediction facet* acts to identify those subpopulations of components that are demonstrated to be at-risk for failure based on the available historical data. The *diagnostic facet* is then a second stage of filtration in which the components within this at-risk subpopulation are categorized based on their present condition using one or more diagnostic tests. Using information from both of these facets, the *economics facet* (or *cost facet*) serves the management function in that the decision to proceed with targeted maintenance is based on the economic benefits resulting from

differences in operating costs and improvements in reliability. Ultimately, through the circular interaction of these facets the *maintenance plan*, which acts on the *aging system*. The *failure prediction* and *diagnostic facets* would then observe the changes produced by the *maintenance plan* and adjust their responses through the *economics facet* to increase, if possible, the resulting benefits.

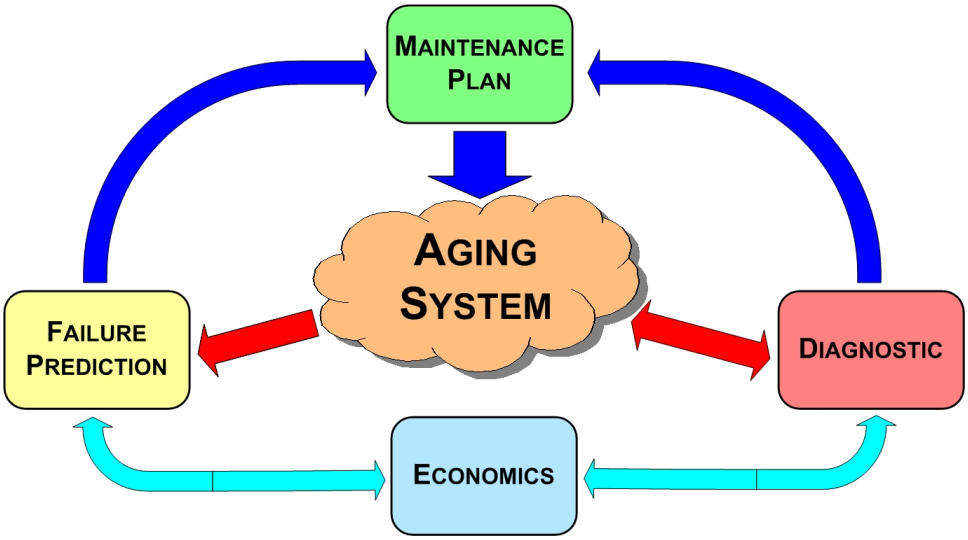
Parts I thru III, describe the details of the failure prediction, diagnostic, and economics facets. The following structure is employed:

- **Part I: Failure Prediction Facet**
    - Chapter 2: Mathematical Models – The detailed derivation of two failure prediction models (Model I and Model II) for use with component historical data.
    - Chapter 3: Point Prediction Results – Results of simulation using two synthesized datasets and one underground cable system field dataset.
    - Chapter 4: Stochastic Modeling – The development of stochastic simulation methodology based on Monte Carlo techniques.
    - Chapter 5: Stochastic Simulation Studies – Results of stochastic simulations from both failure prediction models using the same datasets as in Chapter 4.
- **Part II: Diagnostic Facet**
    - Chapter 6: Diagnostic Accuracy – Mathematics and interpretation of diagnostic accuracy including the key concepts and understanding.
    - Chapter 7: Techniques for Assessing Diagnostic Tests – Development of techniques for the interpretation and assessment of practical diagnostic data.
- **Part III: Economics Facet**

- Chapter 8: The Diagnostic Program – Introduction to the four phases of diagnostic programs: selection, action, generation, and evaluation.
- Chapter 9: The Economics of Diagnostic Programs – Description of the cost functions needed to construct the economic model of diagnostic programs. Also, reviews the interpretation and calculation of "benefit."
- Chapter 10: Economic Simulation Studies – Demonstration of the economic model using case studies based on experience gained through the NEETRAC CDFI project.

- **Part IV: Conclusion**
  - Chapter 11: Summary and Conclusions – Summarizes the research work and the resulting conclusions.

PART I: FAILURE PREDICTION FACET



Part I describes the research conducted to develop a statistical algorithm for the failure prediction facet. This facet is primarily responsible for the processing, analysis, and prediction, of failure trends within a coherent population of components. The process includes an initial prediction phase that is followed by a stochastic simulation based on the Monte Carlo technique. The results from the failure prediction facet are used by the economics facet as a first filtration step in which sub-populations of components are identified as at-risk for failure.

The following four chapters describe the various aspects of the failure prediction facet:

- **Chapter 2: Mathematical Models** – This chapter presents the detailed mathematical derivations of the two Weibull-based failure prediction models, Model I and Model II. Each model allows for the estimation of the number of replacement components needed to achieve a desired failure performance.
- **Chapter 3: Point Prediction Results** – Using the two synthesized datasets and one field dataset, the prediction capabilities of each model are demonstrated. Comparisons between Model I and II help to establish the value of the additional information required by Model I.
- **Chapter 4: Stochastic Modeling with Monte Carlo** – This chapter describes how Monte Carlo techniques may be employed with the failure prediction models to allow for stochastic simulation. Such simulations provide the probabilistic information needed to assess the likelihood for predicted outcomes to occur.
- **Chapter 5: Stochastic Simulation Results** – This chapter presents the results of limited stochastic simulations performed using both models and all available datasets. A variety of failure reduction scenarios are employed and

the resulting differences in the performance of Model I and Model II are clearly shown.

CHAPTER 2: MATHEMATICAL MODELS

This chapter introduces the two failure prediction models, Model I and Model II, that have been developed as part of this research.

2.1 GENERAL HAZARD MODEL

The general mathematical model begins by considering a homogeneous population of devices that are installed at different times during the overall power system's lifetime. If components installed in the same year are said to be part of the same group of components then that group can be thought of a single "vintage" One may then define the evolution of each vintage population. Let $X_{j,k}$ represent the number of components installed in year j remaining in service in year k (assuming the time resolution of the available data is years) where $k \geq j$. Then by definition, the initial population size for any given year y is equal to $X_{y,y}$.

Over time, some of these components will fail and then be replaced by new ones. These failures can be assumed to occur independently and as the total population of components is considered homogeneous then the corresponding times to failure (TTF) will follow the same distribution. These failures will be reflected as a change in the component population as follows: Suppose $X_{1,6}$ includes 30 components and three experience failures in year seven, then $X_{1,7}$ will be equal to 27 and $X_{7,7}$ will include the installation of three replacement components for the failed units that were originally installed in year one. By observing these failures and the corresponding changes in vintage populations one may compute the times to failure for all failed devices and subsequently construct the lifetime distribution.

All lifetime distributions can be described by a *probability density function* (PDF) and a *cumulative distribution function* (CDF) defined as p and P, respectively. The PDF and CDF are related as,

$$P(t) = \int_0^t p(\tau) d\tau. \tag{2.1}$$

These two functions can be combined to give other functions that are useful for prediction purposes. Specifically, one may define the *hazard* function ($h(t)$) as the probability of failure of a component during the next time interval given that it survived to the previous time interval [11]. In reliability, the hazard function is also known as the failure rate [failures/unit time] and is defined mathematically in terms of the PDF and CDF as [20],

$$h(t) = \frac{p(t)}{1 - P(t)}. \tag{2.2}$$

Equation (2.2) represents the probability of failure for a single component as a function of time. However, for a homogeneous population of components the same hazard function will apply to each component individually. Depending on the configuration of the system (series, parallel, or a combination of series and parallel) one may combine the component hazard functions to yield a system hazard function. In the simplest case, a series system experiences a failure if any one component fails. Therefore, the system hazard function ($h_S(t)$) for an X-component system will correspond to the summation of the component hazard functions as,

$$h_S(t) = \sum_{i=1}^{x} h_i(t). \tag{2.3}$$

If all X components are identical then each component would have the same hazard function. Furthermore, if each component is the same age, then the system hazard function simplifies to [28],

$$h_S(t) = \sum_{i=1}^{X} h_i(t) = X\,h(t). \tag{2.4}$$

This definition may be extended to a system that includes components installed in different years by introducing a time shift operator into (2.4) as,

$$h_S(t) = \sum_{i=1}^{n} h_i(t-i) \;,\; t > i \;, \tag{2.5}$$

where,

n = number of component vintages,

h_i = hazard function for each component installed in year i.

Unfortunately, (2.5) only considers the contribution that a single component from each vintage makes to the system hazard function. To include all components in each vintage it is necessary to track the changes in the component population over the entire recorded life of the system. As defined earlier, $X_{j,k}$ represents the number of components from year j remaining in service in year k. Therefore, the overall system hazard function for a series system with components installed over multiple time periods is,

$$h_S(t) = \sum_{i=1}^{n} X_i(t)\,h_i(t-i) \;,\; t > i. \tag{2.6}$$

Equation (2.6) makes no assumptions about the form of the distribution other than it being defined on the interval [0, ∞).

2.2 DISTRIBUTION FITTING

The derivation, thus far, has not assumed a form for the distribution of time to failure data. The construction of any distribution may be approached from two directions: (1) parametric and (2) nonparametric. The former corresponds to fitting in some sense one of the many known distributions including Gaussian (normal), Weibull, exponential, and several others. The latter uses only the data itself to construct the distribution and makes

no assumption as to the shape of that distribution. Both approaches offer advantages and disadvantages, as described in the following limited discussion.

2.2.1 Parametric Distributions

For the case of parametric distribution fitting, the primary advantage is that the behavior and properties of the chosen distribution are already known. As a result, once the necessary parameters for the distribution of interest have been determined they may be used to compute any number of characteristics. In addition, the parametric distributions can be used to overcome issues with the data itself, such as too few samples or unreliable data [30]-[31]. In other words, the parametric distribution may provide additional information that may or may not be available from the data itself.

The primary disadvantage of any parametric distribution fitting is that the chosen distribution may or may not represent the data's true underlying distribution. Therefore, it is imperative that goodness-of-fit tests be performed to determine how likely the data are to fit the particular distribution. Unfortunately, these tests are more useful in proving that the chosen distribution is incorrect. In most cases, a "good" fit is determined through one's own judgment and experience rather than hard numbers. In this case, the most useful tool is that of the probability plot or *q-q* plot [30].

2.2.2 Nonparametric Distributions

The main advantage of using a nonparametric distribution is that no assumptions are made about the distribution being sought. The available data is all that determines the distribution. Techniques such as Kaplan-Meier can be employed to construct the distribution, which, not surprisingly, will fit the data perfectly [32].

Unfortunately, it is far more difficult to extract useful information form nonparametric distributions than it is from their parametric counterparts. Any property of

interest must be estimated from the data itself and many times the data do not contain "sufficient" detail or conditioning to provide that information.

2.2.3 Failure Data Distribution Options

For this research, the decision was made to use a parametric distribution as the basis for the failure prediction facet. Several distributions were considered including Gaussian (normal), Weibull, and time-varying exponential. The last option describes a sliding window type calculation in which the failure rate (λ) is computed over a finite window. The properties of the normal distribution make it the only option that is not appropriate for the desired analysis. This is primarily because of the normal's definition between $-\infty$ and $+\infty$ and its symmetry. Time-varying exponential and Weibull are both useful options. However, given the flexibility and acceptance in the reliability community, the decision was made to utilize the Weibull distribution [30].

2.2.4 Practical Input Data

The decision to utilize a parametric distribution is also strongly influenced by the availability of field data. Parametric distributions can be applied in situations where available data contains only limited information [30]. In the derivation in Section 2.3, the model is based on the availability of time to failure data. Such data requires that both the installation and failure dates of each component are known. For practical planning purposes, it is sufficient to be able to associate the year of installation (vintage) with the year of failure. Unfortunately, most utilities have neglected to maintain records with even this level of detail. A typical utility only records the total numbers of failures and replaced components and does not indicate which vintage populations experienced those failures and replacements.

As asset management strategies are increasingly employed in the electric utility industry planners are realizing that the times to failure are important for addressing

system reliability. As a result, these data are now more often being recorded. However, it will be some time before a sufficiently large amount of data can be amassed.

In light of the currently available data and that data which will become available, this research has led to the development of two failure prediction methodologies. These methods differ primarily in their assumptions regarding the input data. Model I assumes the time to failure data is available for each vintage population while Model II assumes only the annual numbers of installations, replacements, and failures are known.

Sections 2.3 and 2.4 will describe the mathematical derivations of Models I and II, respectively.

2.3 MODEL I - ADVANCED WEIBULL PREDICTION MODEL

This section shows the development of the mathematics behind Model I. It shall include discussions pertaining to the calculation of the model parameters, prediction of failures, and estimation of replacement components.

2.3.1 Estimation of Model Parameters

The estimation of the two Weibull distribution parameters, α and β, begins with the calculation of the times to failure (T) for all installed components. The dataset shall consist of n years of data with known annual failures for each of the resulting n vintage populations of components. Let $f_{j,k}$ represent the failures experienced by vintage population j in the k^{th} year of the dataset ($k > j$). The replacements in each vintage population are also known for each year. As defined in Section 2.2, $X_{j,k}$ represents the number of components installed in year j remaining in service in year k. For discrete components, the difference between $X_{j,k}$ and $X_{j,k+1}$ is the number of failures the j^{th} vintage population experiences in year $k+1$.

Both the failure information (f) and population information (X) may be compactly represented as matrices,

$$f = \begin{bmatrix} 0 & f_{1,2} & f_{1,3} & f_{1,4} & \cdots & \cdots & \cdots & f_{1,n} \\ 0 & 0 & f_{2,3} & f_{2,4} & f_{2,5} & \cdots & \cdots & f_{2,n} \\ 0 & 0 & 0 & f_{3,4} & f_{3,5} & \cdots & \cdots & f_{3,n} \\ 0 & 0 & 0 & 0 & f_{4,5} & \cdots & \cdots & f_{4,n} \\ 0 & 0 & 0 & 0 & 0 & f_{5,6} & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & f_{n-2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & f_{n-1,n} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \textbf{(2.7)}$$

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & X_{1,3} & X_{1,4} & \cdots & \cdots & \cdots & X_{1,n} \\ 0 & X_{2,2} & X_{2,3} & X_{2,4} & X_{2,5} & \cdots & \cdots & X_{2,n} \\ 0 & 0 & X_{3,3} & X_{3,4} & X_{3,5} & \cdots & \cdots & X_{3,n} \\ 0 & 0 & 0 & X_{4,4} & X_{4,5} & \cdots & \cdots & X_{4,n} \\ 0 & 0 & 0 & 0 & X_{5,5} & X_{5,6} & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \ddots & \ddots & X_{n-2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & X_{n-1,n-1} & X_{n-1,n} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_{n,n} \end{bmatrix}. \quad \textbf{(2.8)}$$

Note in (2.7) that the main diagonal of f contains all zero entries while the main diagonal of X in (2.8) is nonzero. In the latter case, the diagonal of X represents the new installations for each vintage population. At the moment of installation, the components begin their first year of aging. Matrix f in (2.7) represents the failures that occur during aging years. In other words, the earliest a failure may occur is during the first year of aging that is represented by $f_{j,j+1}$ for the j^{th} vintage. Those failures that occur between the end of the first year and the end of the second year are represented by $f_{j,j+2}$. This can be generalized as follows: for vintage j, the ages of the components that fail in year m are equal to m-j.

The process of computing the times to failure for all components results in two classes of times: (1) failure times (T) and (2) censor times (S). The failure times come from those components that actually experience a failure before year n of the dataset. The

censor times, on the other hand, come from those components that do not experience a failure before the end of the dataset. These are represented by the n^{th} column of X.

The failure times are computed for each vintage as,

$$T = \left\{ T_k \mid T_k = \sum_{i=1}^{n-k} f_{i,i+k}, \forall k \in \{1,2,...,n-1\} \right\}, \tag{2.9}$$

where,

n = total length of the dataset in years.

Equation (2.10) shows the definition of the censor times set.

$$S = \left\{ S_h \mid S_h = X_{h,n}, \forall h \in \{1,2,...,n\} \right\}. \tag{2.10}$$

Together, (2.9) and (2.10) represent the complete distribution of times to failure. The corresponding CDF may then be fitted using the definition of the Weibull CDF,

$$P(t) = 1 - e^{-(t/\alpha)^{\beta}}, t > 0. \tag{2.11}$$

The Weibull parameters may then be estimated using either least-squares (LS) or maximum likelihood estimation (MLE) approaches. Given the typically large number of censor times, the maximum likelihood technique is preferred since it is better able to adjust the parameters to the censor times as well as to the failure times. The MLE estimates of β and α are shown in (2.12) and (2.13), respectively. Note that numerical techniques (such as those included in statistical software packages) must be employed to solve (2.12). The solution to (2.12) is the MLE estimate of β.

$$\frac{\sum_{i=1}^{M} f_i^{\beta} \ln(f_i)}{\sum_{i=1}^{M} f_i^{\beta}} - \frac{1}{R}\sum_{i=1}^{R} \ln(f_i) - \frac{1}{\beta} = 0, \tag{2.12}$$

$$\hat{\alpha} = \left(\frac{\sum_{i=1}^{M} f_i^{\beta}}{R} \right)^{\frac{1}{\beta}}, \tag{2.13}$$

where,

R = Number of failures,

M = Number of failures (R) + number of censors,

$f_i = i^{\text{th}}$ time to failure or censoring time.

Once the Weibull parameters have been estimated combining them with the known vintage installations (diagonal of X) predictions may be made any number of years into the future. The following section describes the prediction process using this model.

2.3.2 Prediction Using Model I

The prediction process is relatively straightforward once the Weibull parameters have been estimated as described in the previous section. It is based on the use of the hazard function derived in Section 2.2 for a general distribution. Using the definition of the Weibull PDF,

$$p(t) = \beta \cdot \alpha^{-\beta} \cdot t^{\beta-1} e^{-(t/\alpha)^{\beta}}, \quad t > 0, \tag{2.14}$$

and CDF defined in (2.11), the resulting Weibull hazard function can then determined as,

$$h(t) = \frac{p(t)}{1 - P(t)} = \beta \cdot \alpha^{-\beta} \cdot t^{\beta-1}, \quad t > 0, \tag{2.15}$$

where,

$p(t)$ = Weibull probability density function,

$P(t)$ = Weibull cumulative density function,

α = Weibull scale parameter,

β = Weibull shape parameter.

Equation (2.15) assumes all the components are installed in the same year and that only one failure mechanism is present. For a population of components installed over multiple years the resulting model is,

$$h_S(t) = \sum_{i=1}^{n} X_i(t-i)\,\beta_i \alpha_i^{-\beta}(t-i)^{\beta_i - 1} \quad , \; t > i, \tag{2.16}$$

where,

$X_i(t)$ = The number of components installed in year i that are still in service in year t.

α_i, β_i = Weibull parameters for component population installed in year i.

In general, the populations installed in each year (vintages) may be characterized by different Weibull parameters and, as a result, different failure mechanisms. However, in practice there is generally no discernible change in manufacturing or handling that would physically explain these differences. Any difference in the Weibull parameter estimates is more likely due to the MLE process itself. Therefore, it is normal to characterize all vintages using one set of Weibull parameters [30]. This reduces the model to the following form,

$$h_S(t) = \sum_{i=1}^{n} X_i(t-i)\,\beta \alpha^{-\beta}(t-i)^{\beta - 1} \quad , \; t > i, \tag{2.17}$$

where,

α, β = Weibull parameters for total component population.

The only remaining step is to adjust the population in future years by the number of failures that occur in the population. In this model, the assumption is made that the total in service population must remain constant. In other words, the failed components are replaced with the same type of component and that the total population must be functional for the system to operate. The model may also be adjusted to reflect the replacement of a component replacement with a different type of component.

2.3.3 Estimating Replacement Components

The estimation of replacement components is based on injecting additional replacements in the first prediction year to alter the predicted failures for *Y* years into the future. The predictions are maintained by vintage for each year as,

$$F = \begin{bmatrix} F_{1,n+1} & F_{1,n+2} & F_{1,n+3} & \cdots & F_{1,n+Y} \\ F_{2,n+1} & F_{2,n+2} & F_{2,n+3} & \cdots & F_{2,n+Y} \\ F_{3,n+1} & F_{3,n+2} & F_{3,n+3} & \cdots & F_{3,n+Y} \\ F_{4,n+1} & F_{4,n+2} & F_{4,n+3} & \cdots & F_{4,n+Y} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ F_{n,n+1} & F_{n,n+2} & F_{n,n+3} & \cdots & F_{n,n+Y} \\ 0 & F_{n+1,n+2} & F_{n+1,n+3} & \cdots & F_{n+1,n+Y} \\ 0 & 0 & F_{n+2,n+3} & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & F_{n+Y-1,n+Y} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}. \quad \textbf{(2.18)}$$

Rows *n+1* thru *n+Y* constitute the failures that are predicted to occur for the replacement components installed in years *n+1* thru *n+Y-1*. The additional replacements that are to be installed in year *n+1* will also contribute to these failures. In fact, these additional replacements will contribute to the failures for each year following year *n+1* and will themselves cause additional replacements in those years. This is a result of the requirement that the total population must remain constant.

The replacement calculation requires both the initial failure predictions (*F*) and a target failure reduction (F_{Target}). This specifies the number of failures that are to be avoided with the additional replacement components. The calculation process is as follows:

1. Determine the total number of anticipated failures without the additional replacement components as,

$$F_{Init} = \sum_{j=1}^{n+y} \sum_{i=n+1}^{n+Y} F_{j,i} \,. \quad \textbf{(2.19)}$$

2. Calculate the required failure reduction according to,

$$\Delta F = F_{Target} F_{Init} . \quad \textbf{(2.20)}$$

3. Compute the total number of failures that one component installed in year $n+1$ will contribute (including future replacements) over Y years. Denote this as $F_{Penalty}$.
4. Sort the vintage populations based on the number of predicted failures per component in service in year n. In order to replace the minimum number of components, sort from highest failure rate to lowest. The choice of Weibull will cause the order to be from the oldest to the newest vintage. Denote this list as V_{sorted}.
5. Select the first vintage in V_{sorted} , denoted as m, and compute the anticipated reduction in failures according to,

$$f_m = \sum_{i=n+1}^{n+y} f_{m,i} - X_{m,n} F_{Penalty} . \quad \textbf{(2.21)}$$

Note that the net effect of replacing vintage m must include the contribution each replacement component will make to the predicted failures over the next Y years. Define the minimum required replacement components as R_{min}.

6. Compute,

$$\Delta f^{(m)} = \Delta f - f_m . \quad \textbf{(2.22)}$$

There are two possible ways to proceed:

- $\Delta f^{(m)} \leq 0$: Proceed to step 7 and compute the portion of vintage m to replace.
- $\Delta f^{(m)} > 0$: Add $X_{m,n}$ to R_{min} and return to Step 5 but select the $m = m+1$ vintage in V_{sorted}.

7. Determine the portion of the vintage m (R_{part}) to replace in addition to R_{min}. Replacement R_{part} is computed as,

$$R_{part} = \frac{\Delta f^{(m)}}{\frac{f_m}{X_{m,n}} - F_{Penalty}}. \quad \textbf{(2.23)}$$

8. Compute the total replacements (R_{Total}) according to,

$$R_{Total} = R_{\min} + R_{part}. \quad \textbf{(2.24)}$$

9. The predicted failures may then be adjusted by removing all failures from vintages *1* to *m-1* and then removing the failures from the fraction of vintage *m* that was replaced. In addition, for each replacement component, failures according to $F_{Penalty}$ must be added back in to the predicted failures.

In theory, by replacing between the extremes of zero and the entire population of components, one will be able to achieve a near zero failure rate for the Y years. Unfortunately, depending on Y, the replacements will start to contribute failures but at a relatively low rate.

By combining failure prediction with the replacement actions calculation, it is possible to model both the current system trend and the effect of future actions on those trends. The following section describes the same process for Model II.

2.4 MODEL II - BASIC WEIBULL PREDICTION MODEL

Similarly to Section 2.3, this section describes the mathematical derivation of Model II. This description shall include processes for predicting failure trends and estimating the required replacement components needed to achieve a specified failure performance.

2.4.1 Mathematical Derivation

Model II relies on much the same mathematics and shows much the same properties of Model I. However, this model does not have accurate time to failure data available for

use in estimating the Weibull parameters. To overcome this lack of information, the model makes the key assumption that all replacement components are first used on the oldest vintage population with components still in service. The derivation of this model continues from (2.6) using the Weibull form of the hazard function for a single component. After rearranging the terms slightly and assuming a population of identical components installed in the same time period, the basic form shown in (2.25) is reached. Note that this equation is similar to that used in the advanced model.

$$F(t) = X(t) \cdot a \cdot t^{b},\ t > 0, \tag{2.25}$$

where,

F(t) = Estimated number of failures at time *t*,

X(t) = The number of components in service as a function of time,

$a = \beta \cdot \alpha^{\beta}$,

$b = \beta - 1$.

Equation (2.25) relies on, at a minimum, the annual quantities of installed components, replaced or retired components, and failures, *f(t)*, for the total population. As before, the overall failure rate will be the summation of the individual vintage population failure rates as shown in (2.26).

$$F(t) = \sum_{i=1}^{n} X_i(t) \cdot a_i \cdot (t - i)^{b_i},\ t > i. \tag{2.26}$$

If, however, the assumption is made that all vintage populations share the same Weibull parameters then (2.26) may be written for each installation year (or vintage), *i*, by shifting the function in time as,

$$F_i(t) = X_i(t) \cdot a \cdot (t - i)^{b},\ t > i, \tag{2.27}$$

where,

i = Component vintage population number (*i* = 1, 2, 3..., *n*),

$F_i(t)$ = Estimated failures at time t from components of vintage i,

$X_i(t)$ = The number of components in service from vintage i as a function of time.

Summing this function over all vintages yields the estimated failures for the entire population as,

$$F(a,b,g,t) = \sum_{i=1}^{n} F_i(t) = \sum_{i=1}^{n} X_i\left(t\right) \cdot a \cdot (t-i-g)^b,\ t > i+g. \quad \textbf{(2.28)}$$

The goal is to then identify parameters a, b, and g, such that the error between the actual observed failures and the estimated failures is minimized in the least squares sense. The resulting objective function is,

$$\underset{a,b}{Min} \sum_{t=1}^{n} \left(f(t) - F(a,b,g,t)\right)^2. \quad \textbf{(2.29)}$$

The model is completely defined by the component population and the parameters a, b, and g, and can be used to predict future failures.

2.4.2 Prediction Using Model II

Similar to Model I, Model II utilizes a slightly different form of Weibull hazard function to perform its predictions. Using (2.30), predictions may either be made assuming no changes to the component population or assuming the total in-service population remains constant.

$$F(a,b,t) = \sum_{i=1}^{n+Y} X_i\left(t\right) \cdot a \cdot (t-i)^b,\ t > i. \quad \textbf{(2.30)}$$

If the latter assumption is made, then the population matrix will need to be updated at each year with the corresponding change in population. The update procedure is different from that of Model I since the assumption regarding the replacement of the oldest components is still in effect. The annual replacements are computed as,

$$R_T\left(n+m\right) = \sum_{i=1}^{n+m} X_i\left(n+m-1\right) \cdot a \cdot \left(t-i\right)^b. \quad \textbf{(2.31)}$$

These replacements are then applied to the oldest population still in service in year $n+m$. If vintage q corresponds to the oldest vintage still in service in year $n+m$ and $X_q(n+m-1)$ represents the number of components remaining in that population then (2.32) shows the resulting update in population.

$$X_q(n+m) = X_q(n+m-1) - R_T(n+m). \quad \textbf{(2.32)}$$

If $X_q(n+m) \geq 0$, then the population update is complete. However, if $X_q(n+m) < 0$ then $X_q(n+m)$ shold be set to zero and the remaining replacements applied to the next oldest vintage, $q+1$. This process should continue until all $R_T(n+m)$ components have been allocated.

2.4.3 Estimating Replacement Components

Using the failure predictions as a baseline, the number of replacement actions required to reduce these initial estimates to some desired level may then be calculated using the model. Thus far this model assumes the same values for a and b for all vintages. Therefore, the oldest components will have the highest failure rate. This assumption is a consequence of the available historical data and, in practice, may or may not be the case. However, given this assumption and the structure of the Weibull model employed, it is straightforward to alter the failure estimates by performing replacement actions on enough of the oldest components. This effectively eliminates the failures these components would have contributed to the overall population's failures.

At this stage, if replacement actions are performed in the next year then, not surprisingly, the model predicts fewer failures will occur. Depending on the number of replacements, it is possible to alter the failure curve. In principle, the number of replacement components can be anywhere from zero to the total number of components in the system. As a result, it must also be possible for the resulting number of failures to

be between an upper bound (determined by zero replacement scenario) and lower bound of zero (total replacement scenario).

The number of replacement components can be viewed as a control input for the failure rate. However, these variables are related through the model developed. Therefore, the process may be reversed so that the replacement schedule needed to achieve a specified failure level may be computed. Suppose year $n+1$ represents the first prediction year, D_{n+1} the desired failure level for year $n+1$, and F_{n+1} the estimated failures without replacement. It is assumed that $D_{n+1} < F_{n+1}$ as any desired failure level above the initial failure forecast will require zero replacements. The process for computing the required replacements R_{n+1} is as follows:

1. Compute the required reduction in failures according to,

$$\Delta_{n+1} = F_{n+1} - D_{n+1}. \tag{2.33}$$

2. Identify the oldest population (X_m) with components still in service in year $n+1$.
3. Compute the number of failures ($F_{m,n+1}$) population X_m contributes to the total failure rate using (2.26).
4. If $F_{m,n+1} < \Delta_{n+1}$, then additional populations will be required so this entire X_m population will be replaced. Update Δ_{n+1} and R_{n+1} using (2.34) and repeat steps 2-4 for the next oldest population using the updated values,

$$\begin{aligned} \Delta'_{n+1} &= \Delta_{n+1} - F_{m,n+1}, \\ R'_{n+1} &= R_{n+1} + X_m. \end{aligned} \tag{2.34}$$

   If $F_{m,n+1} > \Delta_{n+1}$, then compute the portion of X_m to replace to achieve failure reduction Δ_{n+1} using,

$$R_{n+1} = \frac{\Delta_{n+1}}{a\left(n-m\right)^b}. \tag{2.35}$$

The total required replacements will be the sum of all vintage populations that must be replaced completely with the portion of one population that will require only partial replacement.

If the population requires a constant total number of components then the same steps as those used for Model I must be included to maintain the population.

Either of the two failure prediction models may be employed to generate predictions. However, this has only so far been shown for one year into the future. Section 2.5 describes how these models may be extended farther into the future.

2.5 MULTIYEAR REPLACEMENT CALCULATIONS

The procedures described in Sections 2.3.3 and 2.4.3 may be extended beyond the first prediction year. However, all years $n+2$, $n+3$, ..., $n+k$, would require assumptions for the replacements and resulting failure rates for the preceding prediction years. This constitutes an additional source of uncertainty. For example, computing the number of replacements for year $n+2$ would require knowing what occurs in year $n+1$ in terms of replacements and the resulting failure rate. Obviously, this problem grows rapidly as predictions are made farther and farther into the future. Furthermore, the results become less reliable (the distributions of the predictions would have a wider spread) since they rely more and more on other predictions.

The solution adopted for this research is based on the practical application of resources within a utility. Each utility must maintain an enormous fleet of equipment. In general, the financial resources that would allow for proactive replacement of components as called for by both prediction models are only available when there is a major problem. As a result, asset managers must correct the problem in a short amount of time rather than on a year by year basis. In other words, if the utility experiences an increased failure rate in distribution transformers then resources will be made available to deal with that immediate problem. Such resources will then be devoted to the next major

problem the following year. This means that a year by year replacement program would not be particularly useful since the required resources would disappear within a year or two.

This has led to the implementation of a policy in which all required replacement components are installed in the first year. The added benefit to this scheme is that studies have shown that replacement actions make a larger impact on the system reliability the sooner they are installed. In other words, a replacement population of 100 components installed today will lead to a greater decrease in failure rate for the entire five year time horizon as compared to installing 20 components each year for the next five years.

From the modeling perspective, this practical approach means that the predictions will not need to rely on data that is increasingly composed of prediction data. The only alternative would have required nested stochastic simulations, or convolutions, that would be highly computationally intensive routines and would lead to wide ranging distributions.

2.6 SUMMARY

This chapter has shown the mathematical development of each of the two failure prediction models based on the Weibull hazard function. These models may each be used to generate multiyear failure predictions as well as estimates of the required replacement actions needed to achieve a desired failure performance. The primary difference between the two models is an assumption regarding the level of detail available in the input data. This translates into alternate methods of applying replacement components to the vintage populations. The effect of this assumption shall be investigated in Chapter 3 and Chapter 5.

CHAPTER 3: POINT PREDICTION RESULTS

This chapter presents some sample results of failure predictions and replacement component estimates for both synthesized datasets and field data. These results constitute point predictions as they are solely based on the original datasets themselves. In Chapter 4 and Chapter 5, distributions from Monte Carlo simulation will be used to produce similar types of results in distribution form. These distributions will be the result of repeated calculations using random inputs. Each of these calculations will be the same type that shall be demonstrated in this chapter.

Three datasets will be explored. Two of these, the synthesized datasets, will be modeled using both failure prediction models while the field data will only be modeled using Model II. Appendix A describes the generation of the two synthesized datasets. Sections 3.1 thru 3.3 will address the results themselves while Section 3.4 will present some observations.

A key issue of this investigation is the difference in predictions and replacement component estimates for the same dataset using the two different models. As described in Chapter 2, Model I uses data that includes the ages of the components when they fail while Model II assumes that this data is not available. The following sections will illustrate the importance of this additional information.

3.1 SYNTHESIZED DATASET 1

This section illustrates the performance of each failure prediction model using Dataset 1 (Appendix A) as the input. Table 1 and Table 2 show the calculated model parameters for Model I and Model II, respectively.

Table 1: Model I parameters for Dataset 1.

Model I Parameter	Value
α	30.56
β	1.47

Table 2: Model II parameters for Dataset 1.

Model I Parameter	Value
a	8.8×10^{-3}
b	0.499

Figure 3 shows the fit each model produces to the observed numbers of failures.

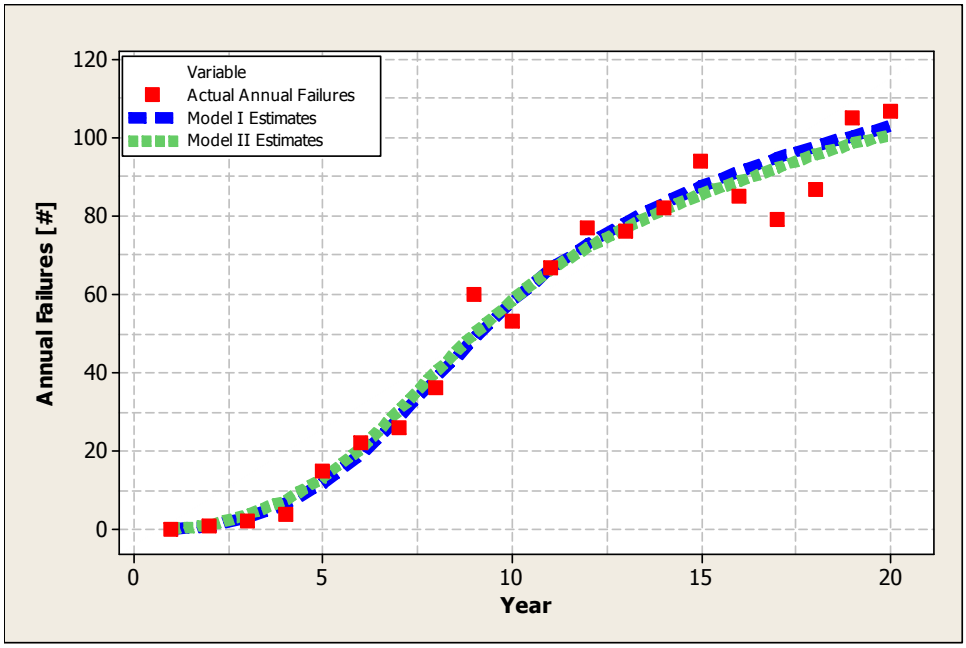


Figure 3: Actual annual failures (■) and failure estimates for Dataset 1 from both Model I (- -) and Model II (---).

A comparison of the quality of the fit each model achieves can be made using the mean squared error (MSE) and mean absolute error (MAE) between the observed failures

and estimates [43]. Table 3 shows the MSE and MAE for each model including their mathematical definitions.

Table 3: Summary of model fit statistics.

Statistic	Model I	Model II
Mean Squared Error $MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2$	35.3	31.7
Mean Absolute Error $MAE = \frac{1}{n}\sum_{i=1}^{n} \lvert e_i \rvert$	4.4	4.3

Based on the statistics in Table 3, it can be said that neither fit is substantially better or worse than the other. The larger MSE for Model I implies that there is more discrepancy between this model and the actual data as compared to Model II. However, when one also considers the MAE, the similar MAE for both models implies that the larger MSE produced by Model I is more the result of a few large errors rather than a general error in the model.

3.1.1 Detailed Model Errors

The distribution of the error with time is shown graphically in Figure 4 where the error is defined as,

$$e = \text{Actual\_Failures} - \text{Estimated\_Failures}. \quad \textbf{(3.1)}$$

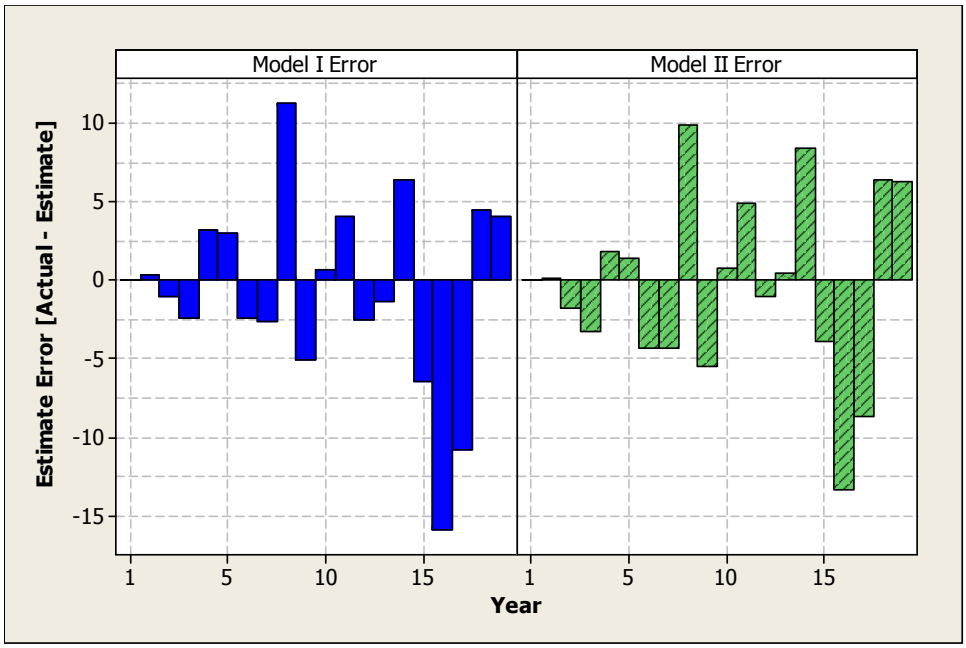


Figure 4: Error between failure estimates and actual failures for Model I (left) and Model II (right). Error is computed as number of failures.

As Figure 4 shows, the error for Model I has a range of [-15.9, 6.4] failures while Model II has a range of [-13.1, 9.3] failures. In general, the two models each over and under estimate the actual number of failures in virtually identical years. This is further evidence that the two models offer very similar performance. Figure 5 shows these same errors as percentages of the observed numbers of failures. Note that Model II shows two large errors in years two and three but these are a result of the very small numbers of failures in those years.

Figure 5: Percent error between failure estimates and actual failures for Model I (left) and Model II (right). Error is computed as a percentage of the observed failures.

Considering the MSE and MAE in terms of percent error yields the results shown in Table 4.

Table 4: Summary of model fit statistics using error percentages.

Statistic	Model I	Model II
Mean Squared Error	482	820
Mean Absolute Error	14.8	15.9

The MSE and MAE using percent error both show that Model I provides a better fit to the data when the relative sizes of the errors are considered. However, again the differences are not terribly large.

3.1.2 Failure Predictions

Figure 6 shows the failure predictions for each model for five years beyond the end of the original dataset. These predictions are summarized in Table 5.

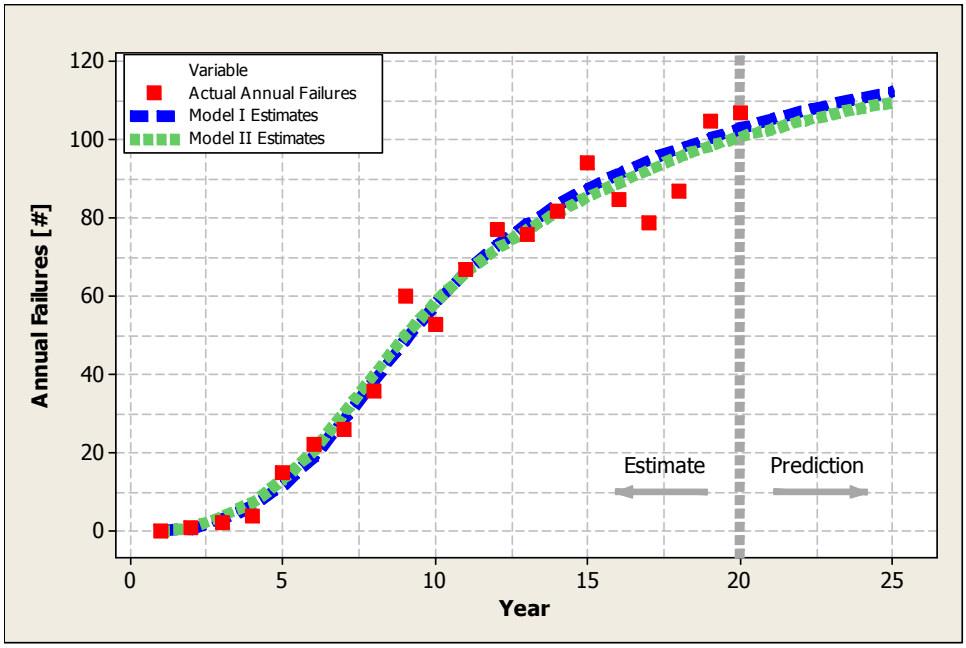


Figure 6: Actual annual failures (■) and failure estimates for Dataset 1 from both Model I (- -) and Model II (---).

Table 5: Comparison of failure predictions made using Model I and Model II

Prediction Year	Model I Prediction	Model II Prediction
1	105.2	102.7
2	107.3	104.8
3	109.2	106.6
4	110.9	108.2
5	112.4	109.6

Table 5 shows that, for this dataset, Model II predicts failures that are approximately 3% lower than the predictions of Model I. This difference is consistent for all the

predictions and is likely the result of the difference in the allocation of replacement components. Such differences in the predictions are, thus, to be expected.

3.1.3 Component Replacement Rates

As described in Sections 2.3 and 2.4, both models allow for the calculation of the number of replacement components needed to achieve a specified failure performance. Figure 7 shows the estimated replacement rates required to produce a reduction in next year's failures of up to 50%.

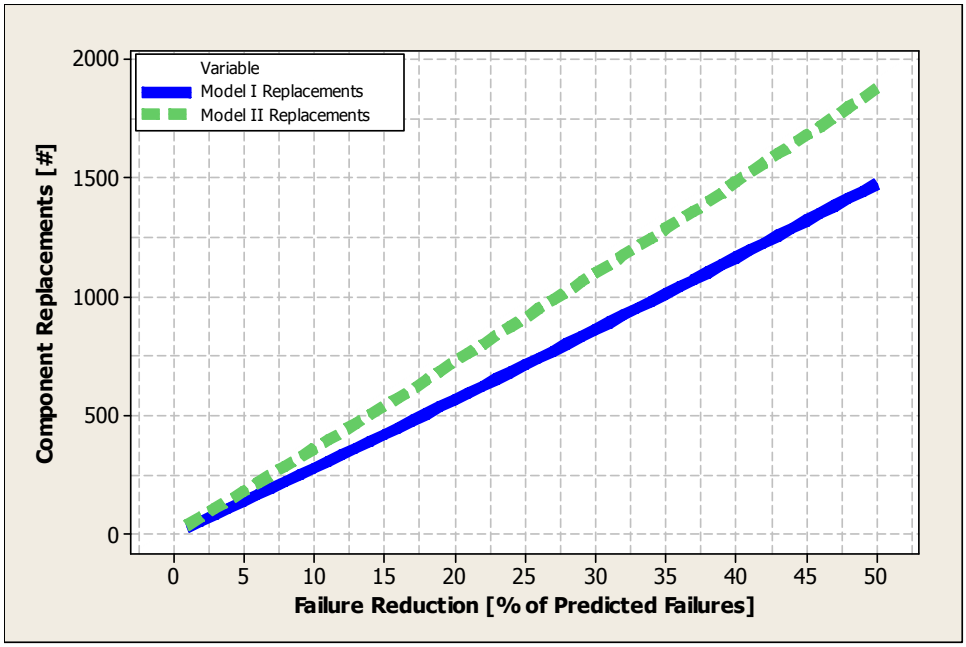


Figure 7: Component replacement estimates based on different failure reduction objectives for next year. Replacement rates for both Model I (—) and Model II (- -) are shown.

At first glance, both curves in Figure 7 appear linear, however, they are not. The linearity is a result of component replacements that are made on the same vintage components. The gradient (slope) changes once each model moves to the next vintage. A more accurate description of these curves is piecewise linear.

As Figure 7 also shows, the required replacement rates for Model II are higher than those of Model I. This effect will be common to all replacement results from Model II. Figure 8 shows the differences between the two models in both actual component replacements and percentage.

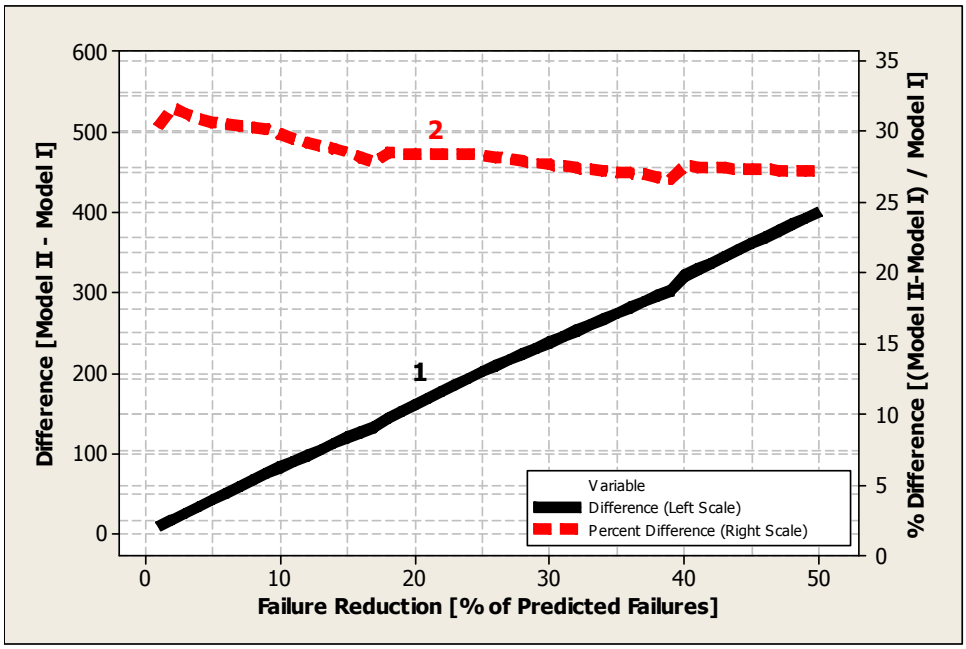


Figure 8: Differences in estimated component replacements for Model I and Model II. Curve 1 (—) refers to left axis while and Curve 2 (- -) refers to right axis.

Figure 8 shows that the difference between the two models increases linearly with the as the number of avoided failures. Furthermore, this difference as a percentage of the Model I replacement rate is in the range of 26-32%. Such a difference leads to significantly higher replacement rates from a Model II estimate for a component population similar to that of Dataset 1. The analysis of Dataset 2 may lead to different results as discussed in the following section.

3.2 SYNTHESIZED DATASET 2

As in Section 3.1, this section will demonstrate the fit achieved by each model using Dataset 2 (Appendix A). Table 6 and Table 7 show the resulting values for each of the model parameters.

Table 6: Model I parameters for Dataset 2.

Model I Parameter	Value
α	20.04
β	2.01

Table 7: Model II parameters for Dataset 2.

Model I Parameter	Value
a	9.3×10^{-3}
b	0.775

Figure 9 shows the actual failure data and the resulting failure estimates from both models.

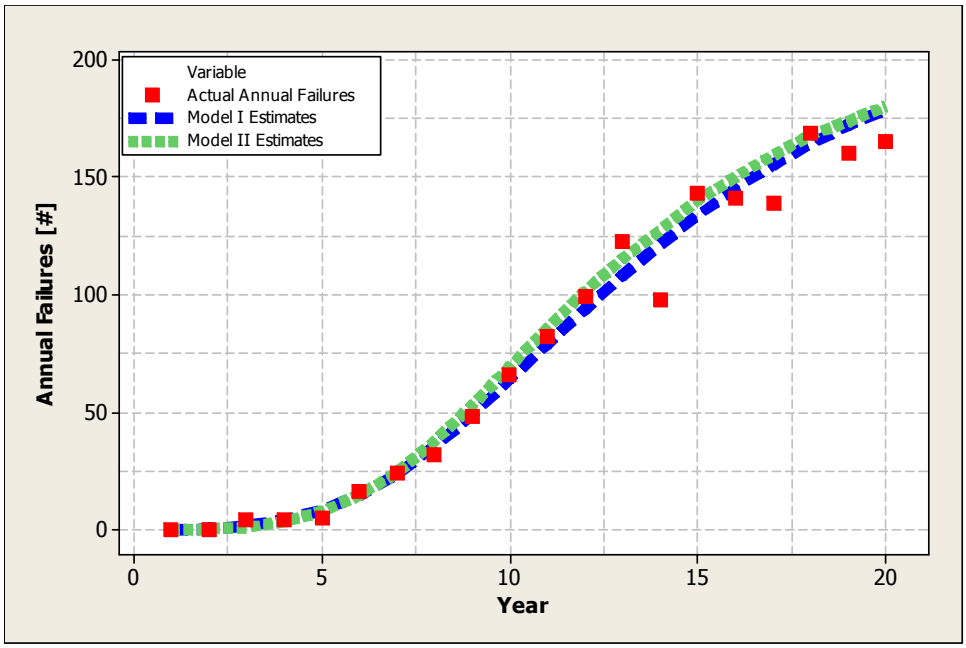


Figure 9: Actual annual failures (■) and failure estimates for Dataset 2 from both Model I (- -) and Model II (---).

Figure 9 demonstrates an interesting behavior in that Model II appears to generally overestimate the failures as compared to Model I. This discrepancy between the two models is shown numerically in the MSE and MAE as shown in Table 8.

Table 8: Summary of model fit statistics for Dataset 2.

Statistic	Model I	Model II
Mean Squared Error	79.04	97.87
Mean Absolute Error	6.07	6.35

Based on the statistics in Table 8, Model I produces a better fit than Model II (lower MSE and MAE). Section 3.2.1 will show the source of the error in Model II.

3.2.1 Detailed Model Errors

The distribution of the error with time is shown graphically in Figure 10.

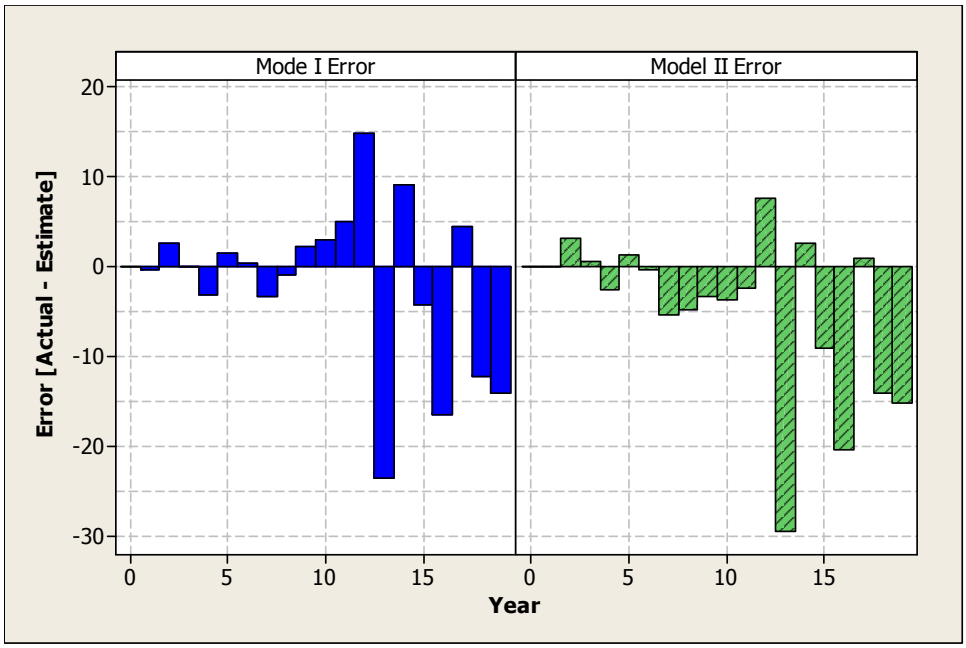


Figure 10: Error between failure estimates and actual failures for Model I (left) and Model II (right). Error is computed as number of failures.

To put these errors in perspective, Figure 11 shows the errors as percentages of the observed numbers of failures for each year.

Figure 11: Percent error between failure estimates and actual failures for Model I (left) and Model II (right). Error is computed as the percentage of actual failures.

Visually, Figure 11 shows that the two models display a similar pattern even though they differ somewhat during years five through ten. During this time interval, Model I tends to underestimate the failures while Model II overestimates them for the same years.

3.2.2 Failure Prediction

Figure 12 shows the predictions for five years for both Model I and Model II. These predictions are shown numerically in Table 9.

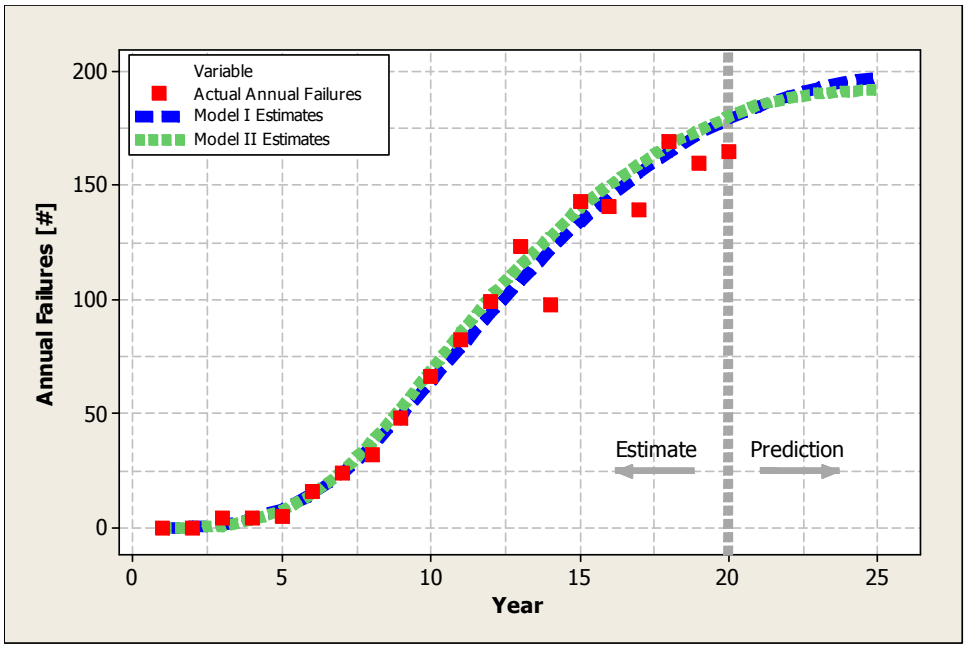


Figure 12: Actual annual failures (■) and failure estimates for Dataset 2 from both Model I (- -) and Model II (---).

Table 9: Comparison of failure predictions made using Model I and Model II.

Prediction Year	Model I Prediction	Model II Prediction
1	184.6	185.2
2	189.1	188.1
3	192.7	190.3
4	195.5	191.4
5	197.5	191.6

Table 9 shows a different behavior than what was seen with Dataset 1. In this case, neither model predicts failures that are either strictly above or below those of the other model. In other words, the prediction curves actually cross between years one and two. In the case of Dataset 1, the Model II predictions were consistently 3% lower than those of Model I and no such crossing point was observed. This difference is likely due to the characteristics of the datasets.

3.2.3 Component Replacements

As with Dataset 1, it is valuable to explore the ability of each model to make replacement action estimates. Figure 13 shows the estimated replacement rates for target failure reductions of 1% to 50%.

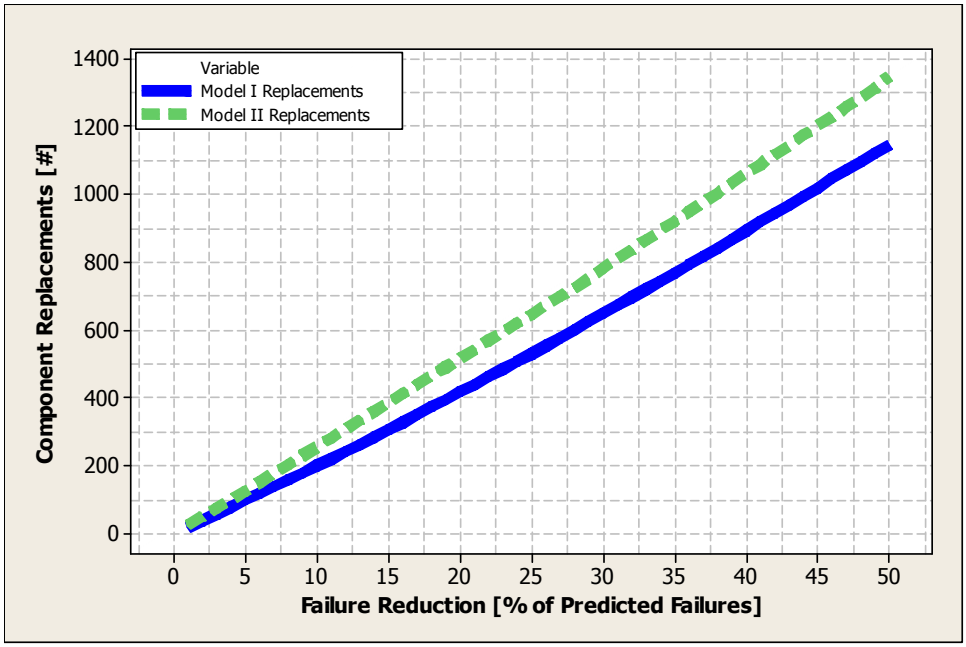


Figure 13: Component replacement estimates based on different failure reduction objectives for next year. Replacement rates for both Model I (—) and Model II (- -) are shown.

The two curves shown in Figure 13 are (as with Dataset 1) each piecewise linear as a result of the transitions between different vintage populations. In this case, however, the curvature is more clearly visible, especially in the estimates from Model I. This is result of the need to replace more vintage populations as compared to Dataset 1.

Figure 13 also shows that Model I again predicts fewer component replacements for each reduction in annual failures. This is consistent with the results from Dataset 1. It reflects the difference in available data that allows Model I to extract more accurately the

evolution of each population. Figure 14 shows the differences between the two models in both actual component replacements and percentage.

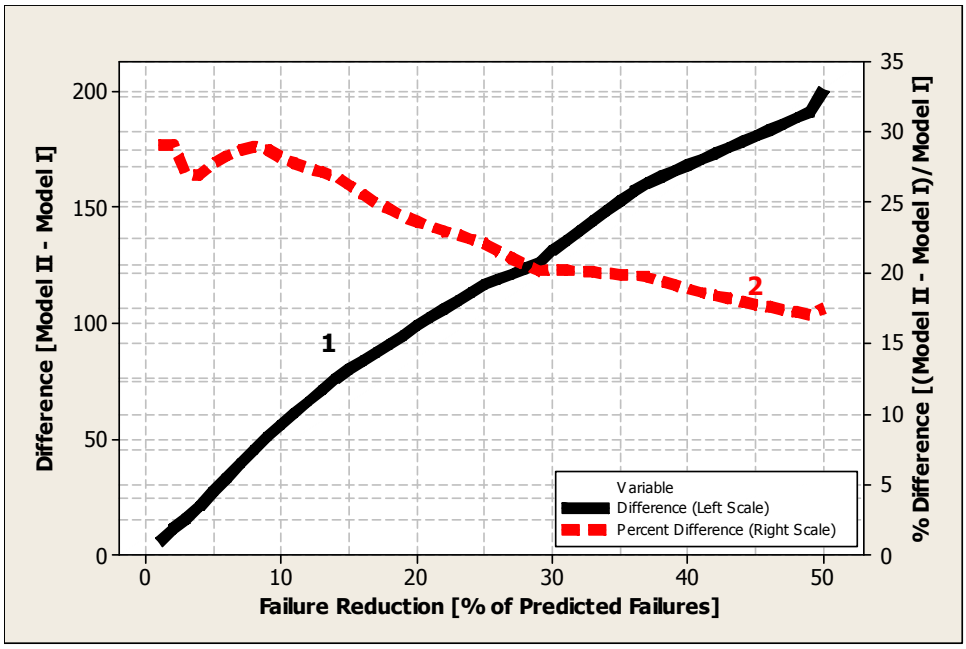


Figure 14: Differences in estimated component replacements for Model I and Model II. Curve 1 (—) refers to left axis while and Curve 2 (- -) refers to right axis.

Figure 14 shows that the difference between the two models increases approximately linearly with the number of avoided failures. The percent difference between the two models is in the range of 17-30% which is somewhat lower and wider distributed than Dataset 1's range of 26-32%. It appears that the difference between the models in both datasets decreases as more replacement components are needed to meet the target failure reduction. However, in both cases Model I estimates fewer replacement components than Model II for the same percent reduction in failures. Surprisingly, this holds regardless of the initial failure predictions since Model I predicts more failures for Dataset 1 and fewer failures for Dataset 2 as compared to Model II.

3.3 FIELD DATA

Appendix A deals with the synthesis of data in order to allow Model I to function. The preference would have been to use real data obtained from a utility. However, as mentioned above, such data are not generally available. In the case of Model II, its reduced data requirements make suitable field data obtainable. This section investigates the performance of Model II using field data for a portion of the underground cable system of a North American utility. Table 10 shows the dataset that will be used in this study.

Table 10: Sample failure and installation data for an underground cable system.

Year	Installs [Miles]	Replacements [Miles]	Total Failures
1963	22.3	10.26	0
1964	43.6	15.13	2
1965	63.3	15.78	0
1966	82.1	10.13	1
1967	104.2	8.9	10
1968	193.5	17.71	6
1969	215.4	16.05	9
1970	328.9	27.95	9
1971	370	60.65	13
1972	416.8	60.48	10
1973	452.7	42.88	17
1974	509.9	29.03	25
1975	437.4	16.54	28
1976	394.6	19.27	32
1977	61.3	4.44	48
1978	16.7	2.95	30
1979	15.5	0.94	43
1980	0	0	45
1981	0	0	57
1982	0	0	53
1983	0	0	68
1984	0	0	75
1985	0	0	67
1986	0	0	81
1987	0	0	88
1988	0	0	100
1989	0	0	84
1990	0	0	127
1991	0	0	154
1992	0	0	139
1993	0	0	137
1994	0	0	156
1995	0	0	151

3.3.1 Initial Observations

Plotting the annual failures from Table 10 shows that this utility has seen a definite decrease in its underground cable system reliability (see Figure 15) as the failure rate clearly increases over time.

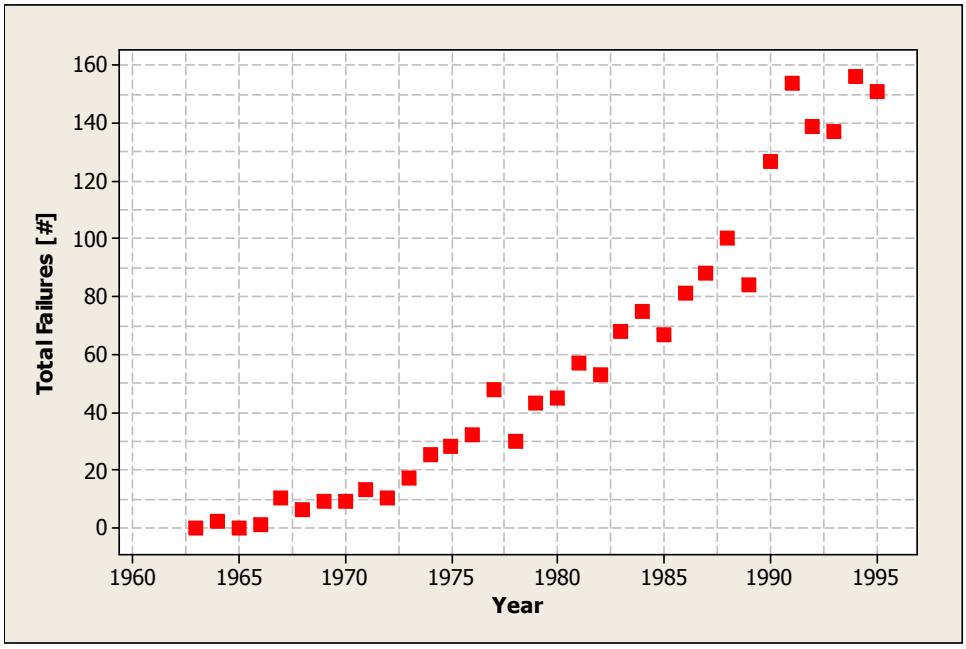


Figure 15: Total number of failures per year for the data shown in Table 10.

On the other hand, Figure 15 is not adjusted to show the effects of the population's growth over time since 1963. Figure 16 shows the failure rate [failures/100 miles] as a function of year. From this figure, it appears that the utility experienced an unusually high failure rate during the first four years of use of this type of cable. However, following this initial period, the failure rate has steadily increased to approximately four [failures/100 Miles/year].

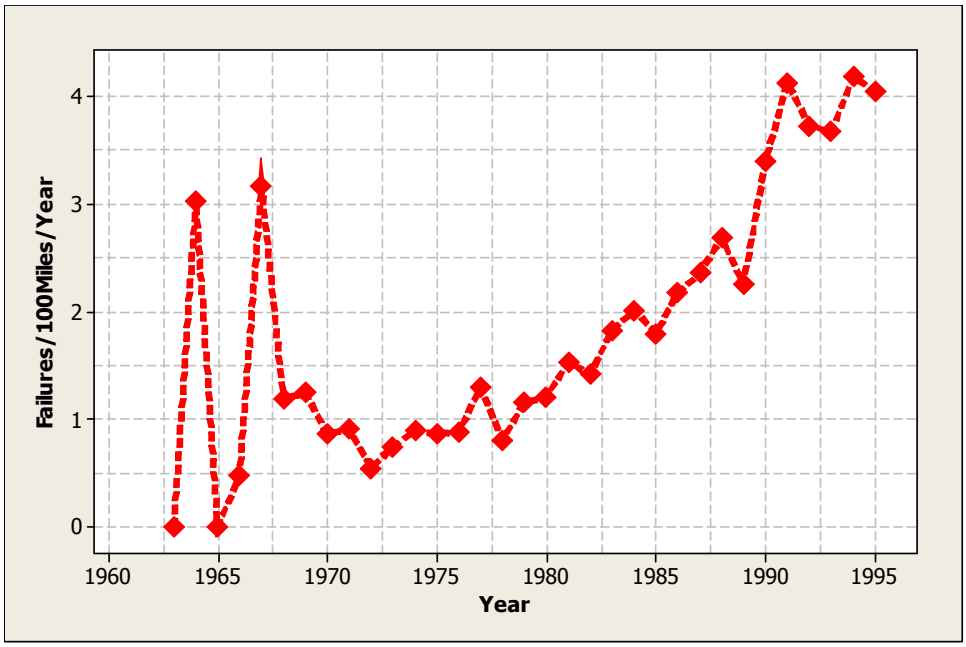


Figure 16: Observed failure rate [failures/100 miles/year] adjusted for the total population of cable in service each year.

Since Model II utilizes a parametric distribution to fit the failure data, it is necessary to verify whether or not the data are consistent with the chosen Weibull distribution. This is accomplished through the use of a probability plot. In such a plot, the data are considered to be consistent with a given distribution if the data points lie close (parallel) to the diagonal line shown and are generally within the confidence bounds shown on the plot. Figure 17 shows the probability plot for the annual failures with Weibull, normal, exponential, and logistic distributions.

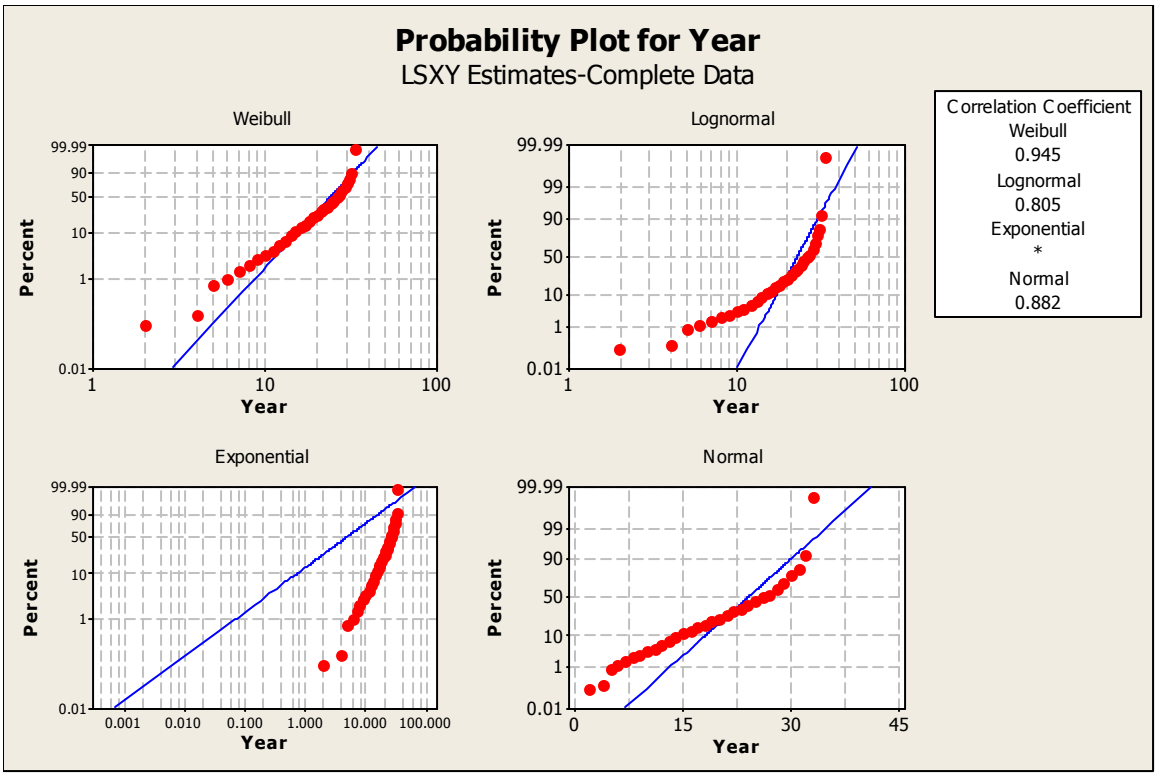


Figure 17: Probability plot of annual failure data using several parametric distributions including Weibull (upper left), normal (upper right), exponential (lower left), and logistic (lower right) distributions.

According to Figure 17, the data are best modeled by a Weibull distribution. However, this may be determined numerically using the correlation coefficient. The closer the correlation coefficient is to one, the better the fit provided by the distribution. Note that in each of the cases the number of samples and corresponding correlation coefficient give the correlations a 0.1% probability of occurring randomly.

Table 11 summarizes these values.

Table 11: Summary of Anderson-Darling statistics for distribution fits shown in Figure 17.

Distribution	Correlation Coefficient
Weibull	0.945
Normal	0.882
Exponential	*
Lognormal	0.805

Table 11 shows that among these distributions the Weibull distribution provides the best fit for the available failure data.

3.3.2 Model Fitting

As the Weibull distribution has been shown to be a reasonable choice for modeling the failure trends in the data shown in Table 10, the data may now be used in the failure prediction model derived in Section 2.4. As mentioned, the ages of each component are unknown at the time of failure and so it is assumed that the oldest components fail and are replaced first. Using the full dataset (i.e., all years of data) the three model parameters may be determined as shown in Table 12.

Table 12: Failure Prediction facet model parameters for data shown in Table 10.

Parameter	Value
a	1.11×10^{-3}
b	1.15

Using these two parameters the curve fit shown in Figure 18 is obtained.

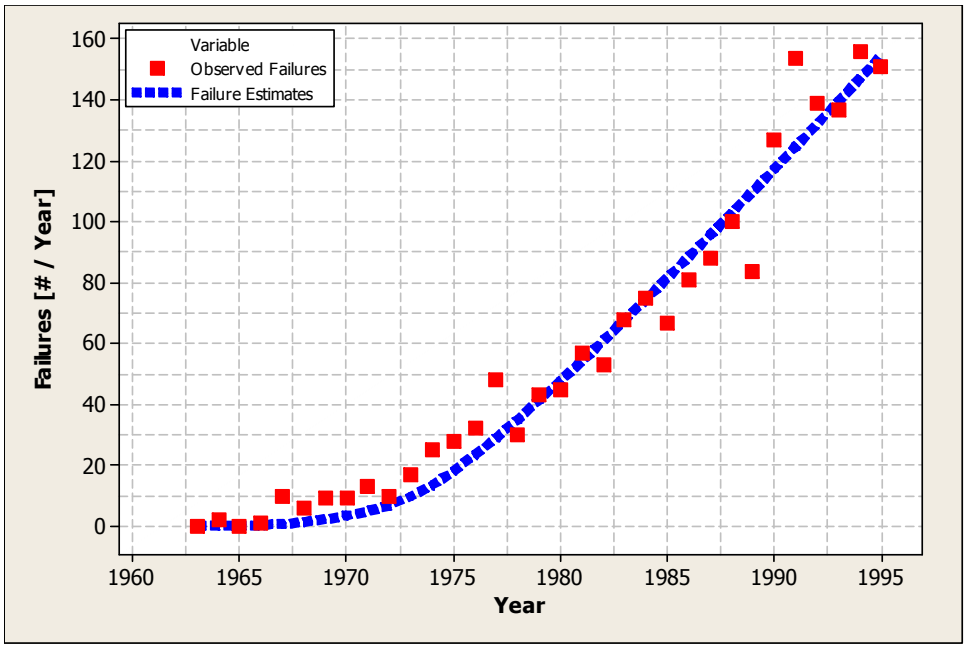


Figure 18: Observed annual failures (■) as shown in Figure 15 and failure estimates computed using Model II (---).

The errors between the estimated failure curve and the actual observed failures are shown in Figure 19 as percentages of the observed annual failures.

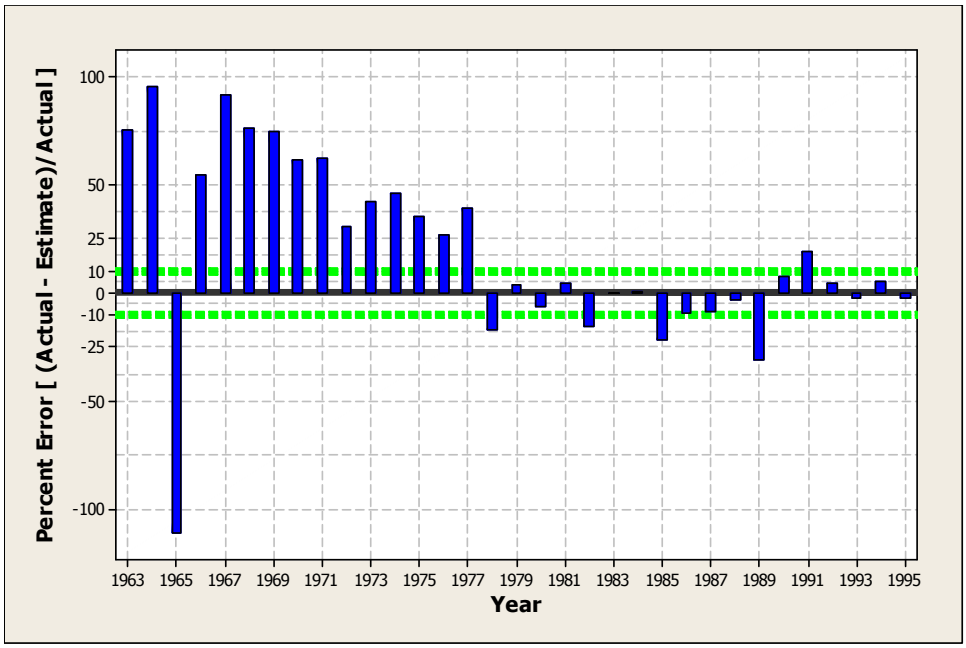


Figure 19: Percent error between actual observed failures and estimated failures.

Figure 19 shows that the estimates improve substantially past 1977 (15 years of data). Up to that point, most estimates are well below the observed numbers of failures, indicating that the *b* parameter is likely underestimated. This in turn implies that the components are, in reality, aging faster than the model estimates.

3.3.3 Verification Using Evolving Window of Failure Data

This section describes one method of validating Model II (and its assumption) known as evolving window. The process operates on the input data, essentially using longer and longer subsets of the data shown in Table 10. For example, one may use all data up to 1980 to predict failures for year 1981. This prediction can then be compared to the observed number of failures. This process may also be used to examine the convergence characteristics of the model parameters, *a* and *b*. This is done by noting that convergence is achieved when the values of the parameters obtained from the subsets are within some small ε of their final values. Figure 20 and Figure 21 show the convergence of *a* and *b*,

respectively. According to these figures, both *a* and *b* converge once data is included through 1991 (year 28). This illustrates the Bayesian paradigm in which new information is incorporated into the calculation as it becomes available.

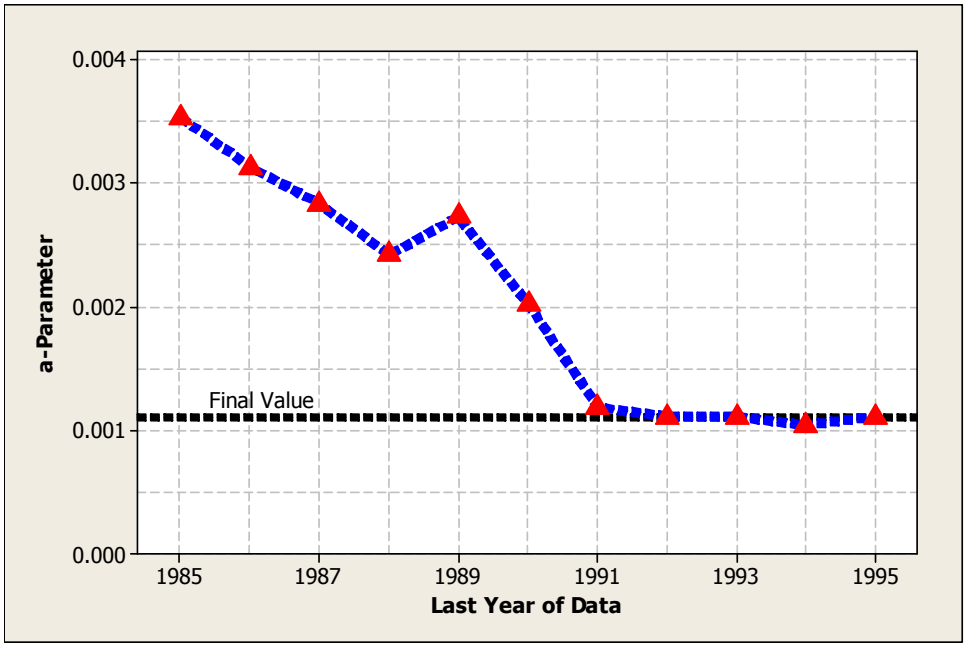


Figure 20: a-Parameter value versus length of dataset. Computed using failure data truncated at year shown on x-axis.

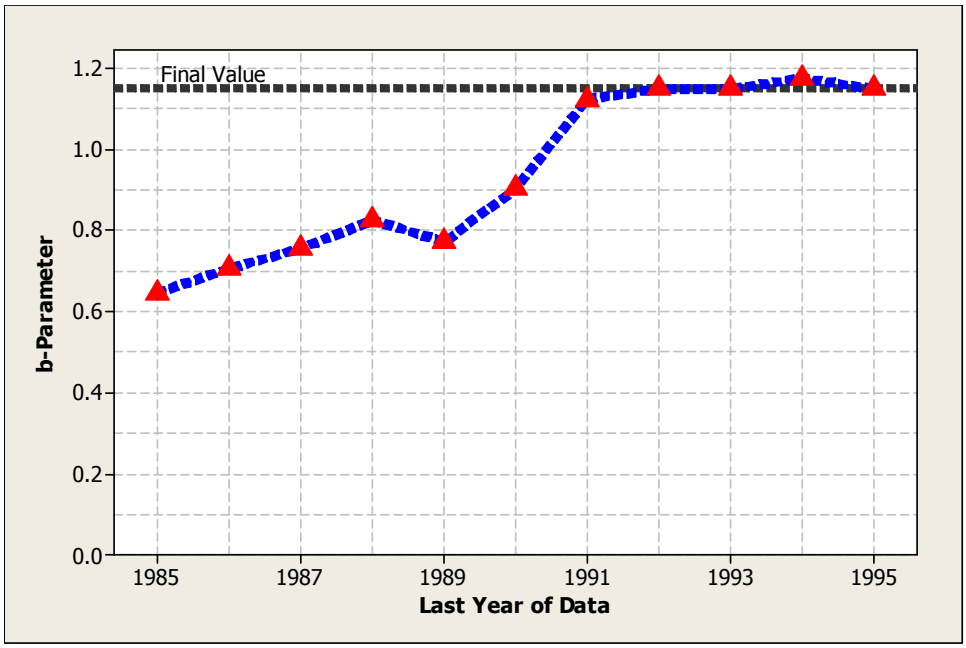


Figure 21: b-Parameter value versus length of dataset. Computed using failure data truncated at year shown on x-axis.

Figure 22 shows the observed failures and predicted failures based on data up through the previous year.

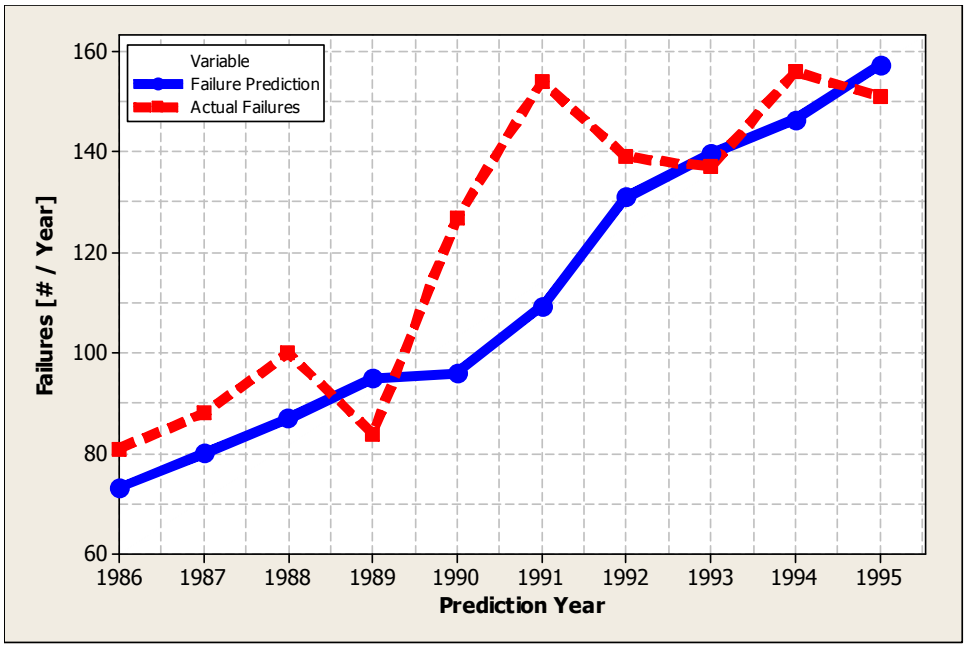


Figure 22: Failure prediction based on data thru previous year and actual failures versus year.

Figure 23 better illustrates the differences between predicted failures and actual failures. As in the case of the model parameters, *a* and *b*, the errors reduce significantly once the dataset used for prediction includes 1991 data. For 1992 and on, the errors are all less than 6% in magnitude.

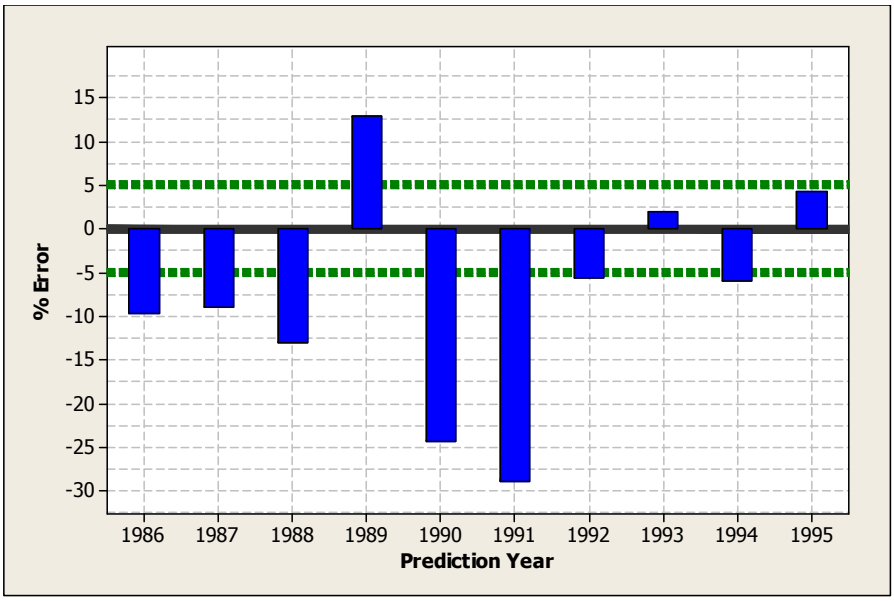


Figure 23: Error between failure prediction and observed failures as percentage of observed number of failures.

Based on these results and those presented in Sections 4.3.1 and 4.3.2, the estimates generated by Model II appear to be valid when field data is used. The primary caveat to this is the amount of available data. As Figure 20 and Figure 21 show, several years worth of data are required before the model parameters can be identified with a high degree of confidence. This is very much dependent on the properties of the dataset itself. Fortunately, datasets that are more consistent (such as the synthesized datasets described in Chapter 3) need not include as many years worth of data as the dataset used in this section.

3.4 OBSERVATIONS

Chapter 3 thus far has investigated the performance of both failure prediction models for estimating future failures and the number of replacement components needed to alter those predictions. This comparison has been made exclusively using synthesized data that is characterized by known Weibull parameters. This data has been needed as a result of

the lack of such data in the industry. One of the objectives for developing Model I was to assess the impact of additional information on the predictions. In this case, the additional information consists of the association of each failure with a particular component vintage. In other words, the age information for each failed component is known.

Based on the results from Sections 3.1.3 and 3.2.3, the advantages in terms of replacement components are clear when the age information is known (Model I). When that data is extracted from the dataset there is a clear increase in the estimates for component replacements. From the study conducted here, this difference can be as much as 32% which would have a significant impact on a utility's ability to follow the replacement schedule of Model II. On the other hand, the results from Model II show that useful predictions can be made in cases where the age information is not available. Clearly, though, there is significant value in beginning to maintain records that would allow the use of Model I.

The field data results presented in Section 3.3 show that the choice of the Weibull distribution is valid for components such as underground cables.

3.5 SUMMARY

This chapter has presented failure predictions and estimated component replacements for a variety of datasets and scenarios. Bear in mind that these predictions and estimates represent only expectations of the output distributions. In the cases where comparison can be made between the two models, Model I generally produces estimates of replacement components that are up to 32% less than those of Model II. Equally, Model II's limited data do not seem to hinder the model as much as might be expected given the severity of the loss of information. Clearly, the preferred model is Model I but in the absence of suitable data, Model II still performs reasonably well.

CHAPTER 4: STOCHASTIC MODELING

Thus far, the focus of Part I has been the derivation and demonstration of both failure prediction models using deterministic examples. The models are, in reality, both probabilistic in nature. However, the results presented have not been used to illustrate their probabilistic nature. This section will examine the use of Monte Carlo techniques for facilitating the stochastic simulation process that is ultimately needed to achieve the goal of the failure prediction facet. The focus will be on the formulation of the stochastic simulation starting with Monte Carlo technique basics and then exploring the details of the chosen implementation.

4.1 MONTE CARLO TECHNIQUE BASICS

Monte Carlo techniques represent a relatively straightforward way of addressing the problem of uncertainty in prediction problems. This uncertainty arises from the fact that the objective of prediction is to determine the outcome of an event that has yet to occur. The fundamental flaw of any such prediction is that it may either be right or wrong. To be of value, predictions must include a probability that provides the likelihood that the specified outcome will occur. Monte Carlo techniques allow one to determine such a probability by inputting random occurrences of input data into the prediction algorithm and then observing the output. When enough of these repeated calculations are performed, all the outputs become probability distributions instead of single values [33]-[35]. The distributions may then be analyzed to extract the confidence intervals thus giving the predictions their respective probabilities of occurrence.

A confidence interval is simply a non-parametric method of specifying the variability or range in a random variable of interest. In the case of prediction, these intervals are

more often termed as prediction intervals. A two-sided prediction interval of (1-α) percent will correspond to a fraction of the samples from a distribution as,

$$\frac{1-\alpha}{100}\times N, \tag{4.1}$$

where,

N = the number Monte Carlo simulations performed.

These intervals provide a measure of how much uncertainty results from the prediction process. A narrow output variable confidence range resulting after significant uncertainty was included in the inputs would represent a well behaved process (although one must be careful not to construct a model that removes uncertainty without evidence for doing so). On the other hand, if wide confidence intervals result from small uncertainties in the inputs that then the process could be considered highly unpredictable. In the first case, one could be relatively confident in the predictions while the latter case would be far less trustworthy. Alternatively, one would be more likely to act on predictions in the first case while less likely to do so in the second case. This is the usefulness of Monte Carlo simulation studies.

The Monte Carlo process is a three step procedure:

1. Define realistic uncertainties in the input variables (this is the most challenging aspect).
2. Compute the outputs of the model for a reasonable number of randomized inputs.
3. Analyze the output distributions in terms of confidence intervals.

As the steps above allude to, the main issues to consider in employing Monte Carlo techniques are:

- How to generate the random datasets from observations?
- How many simulations should be performed?

These questions are addressed in Sections 4.2 and 4.3.

4.2 RANDOM DATASET GENERATION

A number of methods exist for obtaining the randomized inputs for a Monte Carlo simulation. The following techniques have been examined as part of this research:

- Bootstrapping,
- Random sampling of parametric distributions.

Sections 4.2.1 and 4.2.2 discuss these techniques in detail.

4.2.1 Bootstrapping

The Bootstrapping technique has a variety of uses beyond those of stochastic simulation. Uses of Bootstrapping include the testing of classification algorithms, neural controllers, and other pattern recognition methods.

The technique itself makes use of the available "real" input data or the differences between model estimates and this "real" data. Using this set of data, one constructs the random datasets by randomly combining different elements from this set. Consider the following example: Suppose, estimates have been made of failures for an arbitrary dataset that spans 10 years. Denote the input failure data as $f(t)$ and the resulting estimates as $F(t)$ as shown in Table 13.

Table 13: Example dataset and model fit.

Year	$f(t)$	$F(t)$	$f(t) - F(t)$
1	0	0	0
2	2	1	1
3	5	3	2
4	7	7	0
5	9	12	-3
6	13	16	-3
7	18	20	-2
8	26	30	-4
9	29	33	-4
10	35	38	-3

The set ε is defined to be the set of errors resulting from the difference between $f(t)$ and $F(t)$ for $0 < t \leq n$ as,

$$\varepsilon = \left\{ f(t) - F(t), for\ \forall t \in [1, n] \right\}. \quad \textbf{(4.2)}$$

Each of the N random datasets, $\zeta^{(j)}$, is then constructed by randomly selecting elements from ε and combining with $F(t)$ as,

$$\zeta^{(j)} = F(t) + \kappa(t), \quad \forall t \in [1, n], \quad \textbf{(4.3)}$$

where,

$\kappa(t)$ = Function corresponding to the uniform random selection of an element from ε for each value of t,

$j = 1, 2, 3, ..., N$.

This process is illustrated in Table 14.

Table 14: Example dataset with one random dataset generated using bootstrapping.

Year	$F(t)$	$f(t) - F(t)$	$\zeta^{(j)}$
1	0	0	0
2	1	1	3
3	3	2	0
4	7	0	4
5	12	-3	8
6	16	-3	14
7	20	-2	16
8	30	-4	26
9	33	-4	34
10	38	-3	40

The above example illustrates one possible implementation of bootstrapping. As mentioned, there are a number of variations but the basic concept of sampling from a defined set is common to all of them.

4.2.2 Random Sampling

The random sampling technique has also been explored as part of this research. In fact, this technique is used as part of both models for generating the random datasets for Monte Carlo simulation. In its most basic form, this technique casts each input variable as a distribution (either non-parametric or parametric) and then extracts from that distribution one value for each random dataset. This process is repeated for each input variable. It is best to choose distributions that are physically meaningful and are representative of how the inputs might behave in reality. This will greatly improve the usefulness of the resulting outputs.

In addition, one must be cautious not to generate values for the inputs that are impossible to obtain in reality. For example, if generating random times to failure then the chosen distribution must not provide a negative value as this would be physically impossible. The most common example of such a distribution is the normal distribution.

4.2.2.1 Multiple Distributions

Using Model II as an example, the following implementation of random sampling is used: Consider the Weibull distribution parameters obtained using Model II. The random datasets for Monte Carlo simulation are generated by randomizing the annual numbers of failures (F_T). The process begins with the estimates obtained from the model ($F(t)$) and the corresponding shape parameter, β. For every estimated value of $F(t)$, a Weibull distribution is generated using that value as the mean of the distribution. The shape parameter shall remain unchanged for each distribution. These n distributions are then formed by computing a new scale parameter for every t using,

$$\alpha_i = \frac{F(i)}{\Gamma\left(1+\frac{1}{\beta}\right)}, \tag{4.4}$$

where,

$\beta = b+1$,

α_i = Weibull scale parameter for the i^{th} failure data point,

$\Gamma(\,)$ = Gamma function defined as $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$.

For each i, N random numbers are generated using $Weibull(\alpha_i, \beta)$. Each Monte Carlo dataset is constructed by selecting one of the N random numbers from each of the n distributions.

4.2.2.2 Single Distribution

An alternative approach, as used in Model I, is to generate a time to failure for each component. Then, simply count the number of failures for each year and then exclude the times to failure that are greater than the maximum age the vintage population could obtain. For example, in the case of a 20 year dataset, the maximum age of the components installed in year one is 19 years. Similarly, those installed in year two could

achieve at most 18 years of age. For the case of year one, all times to failure longer than 19 years would not be counted as failures in the 20 year long dataset.

Clearly, there are a number of methods for generating randomized datasets. The key is to choose a method that as accurately as possible portrays the uncertainty that can arise in the input variables. The chosen method has a tremendous impact on the resulting output distributions and so should be chosen carefully.

4.3 DETERMINING THE APPROPRIATE NUMBER OF SIMULATIONS

The length of a Monte Carlo simulation is also an important issue. Unfortunately, there is no way to determine upfront how many datasets to use since it is highly dependent on the model and randomization technique. Generally speaking, the more uncertainty present in the input data the more simulations are needed to be able to observe the effects on the outputs. This means that the simulation should produce enough data points to be able to construct each output variable's PDF. As a guideline, one can examine the output distributions as the simulation progresses to observe whether or not the distribution is sufficiently defined. Such a check is based on one's own judgment. If time is not a constraint, then it is best to continue the simulations beyond this point. The following criteria may assist is deciding whether or not the simulation has produced enough data points:

- Confidence Intervals – Are the confidence intervals stable within an incremental number of simulations? If so, then continuing the simulation will not likely provide additional information.
- Distribution Parameters – If a parametric distribution may be fitted to the output data, do the estimated parameters for that distribution remain stable? Do the confidence limits on these estimates remain constant?
- Histogram – Do the histograms of the output variables appear smooth?

4.4 SUMMARY

As this chapter describes, Monte Carlo techniques in general require a great deal of judgment to execute. The mechanics of Monte Carlo are straightforward. However, the decisions made by the researcher are vital to its performance and value. The following chapter illustrates the results obtained from Monte Carlo simulations using both failure prediction models and all datasets discussed thus far.

CHAPTER 5: STOCHASTIC SIMULATION STUDIES

This chapter investigates the performance of both Model I and Model II using Monte Carlo techniques. The simulation results presented in this section are based on the synthesized datasets described in Appendix A and on the field data of Section 3.3. The investigation will illustrate the following:

- Fewer replacement components will be needed in Model I to produce the same reduction in future failure rates as compared to Model II.
- The output distributions will all be normally distributed even though all randomization is based on Weibull distributions.
- Distributions for estimated replacement components will be narrower (smaller standard deviation) for Model I.

For each dataset, a series of simulations are presented in which the goal reduction in failures is defined as 5%, 10%, and 20%, of the total predicted failures for a five year time horizon. In the case of Dataset 1 and Dataset 2, the simulations are performed using both failure prediction models to provide a means of comparing the two models. On the other hand, the field dataset will be demonstrated only with Model II. Sections 5.2 thru 5.4 present the results of these simulations. The following section, Section 5.1, begins with a description of the graphical tools that will be used to present the results of each Monte Carlo simulation.

5.1 VISUALIZING MONTE CARLO RESULTS

There are a number of ways to represent the kinds of results that are inherent in Monte Carlo simulations. As mentioned above, all output variables of the Monte Carlo routines will be approximations of distributions. For example, in the cases that will be illustrated in the next several sections each output variable will consist of 1000 different

values. These data points form a distribution that may be represented as a histogram (as shown in Figure 24) or as a box plot (as shown in Figure 25). The box plot format is a far more compact representation than the histogram as it allows one to compare a greater number of variables to one another in a single graph. This format will be the primary method used to represent Monte Carlo results.

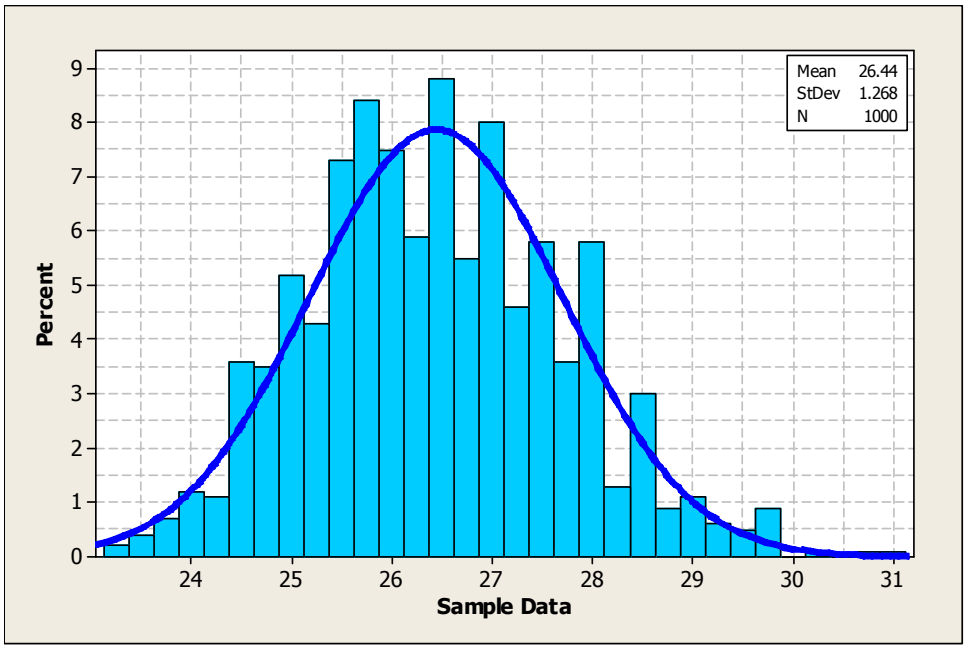


Figure 24: Sample histogram of Monte Carlo data. The x-axis represents the different values of the data while the bars represent the percentage of the data with that value. The line is a normal distribution that is fitted to the data.

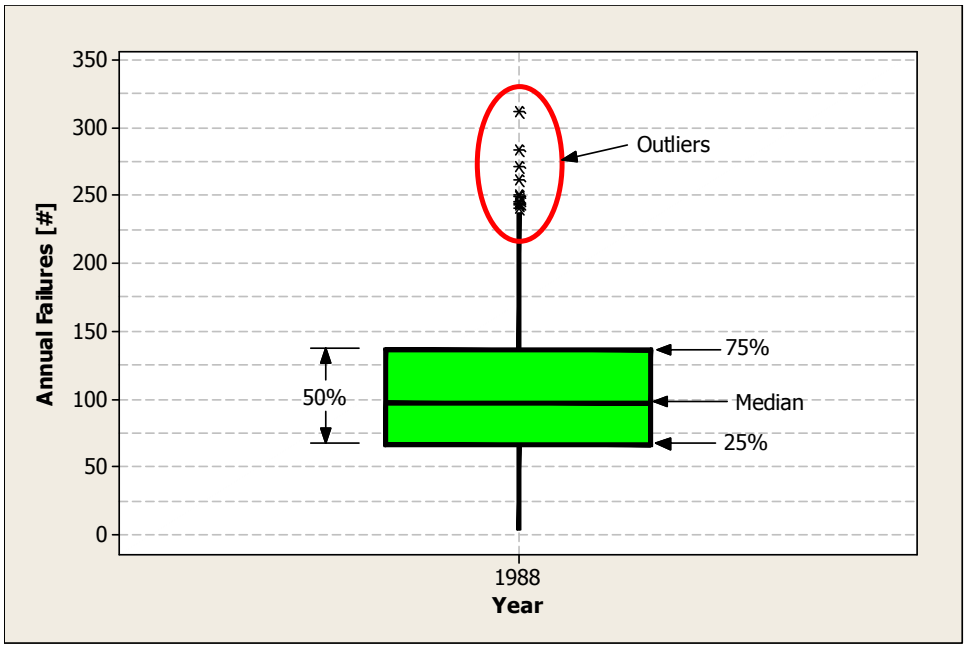


Figure 25: Box plot showing sample data. The box corresponds to the middle 50% of the data while the line through its center is the median. The lines and ticks represent the outlier data.

Yet another method of illustrating the Monte Carlo results is in the form of an upper confidence bound plot. This is most applicable to the distributions of estimated replacement actions. In this case, the upper confidence bound shows the number of replacement actions needed to guarantee with a certain confidence that the failure performance will be either met or surpassed. Figure 26 shows an example of such a curve. For a confidence level of 50%, the maximum number of replacement actions needed is 270 to guarantee that the failure reduction is at least 5%.

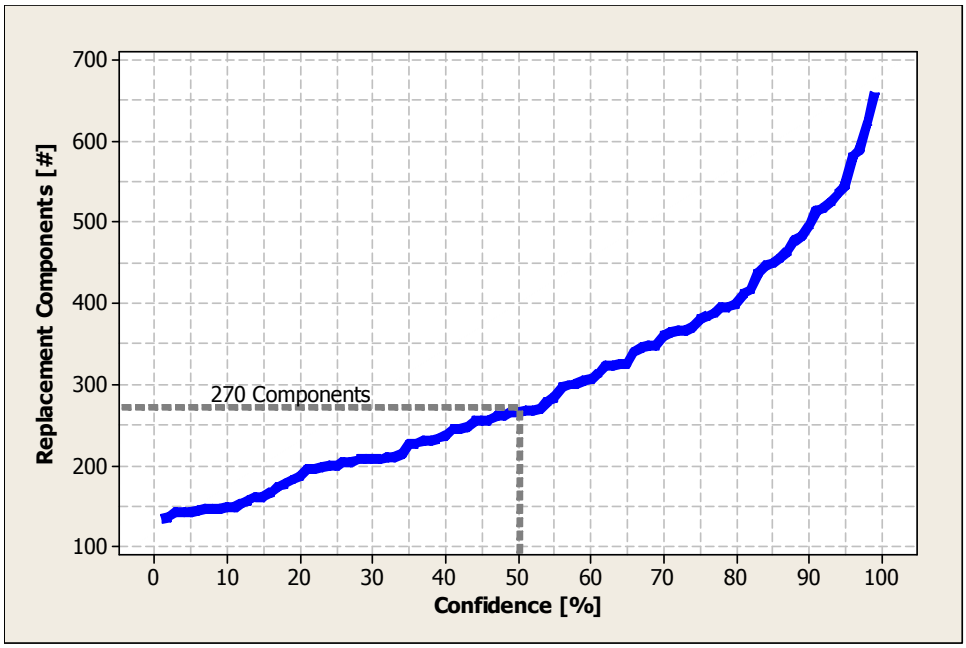


Figure 26: Sample upper confidence plot showing the replacement components needed to guarantee that next year's failure performance meets or exceeds the target level with the specified confidence level.

Sections 5.2 thru 5.4 review the Monte Carlo simulation results for each of the datasets discussed in Chapter 4.

5.2 SYNTHESIZED DATASET 1

This section examines the results of Monte Carlo simulation for Dataset 1 (Appendix A) using both Model I and Model II. The primary interest in both this section and Section 5.3 is the comparison of results from the two models.

5.2.1 Simulation Setup

The Monte Carlo simulation parameters used for Dataset 1 are shown in Table 15. The randomization procedure employed in both models is identical and is based on the random sampling technique described in Section 4.2.2.

Table 15: Monte Carlo simulation parameters for each failure prediction model.

Parameter		Model I	Model II
Simulations [#]		1000	
Prediction Years		5	
Target Failure Reductions		5%, 10%, and 20%	
Dataset Generation Weibull Parameters	α	30.57	30.86
	β	1.47	1.50

5.2.2 Estimated Replacement Actions

Figure 27 shows an example of the distribution produced by each failure prediction model for estimated replacement actions.

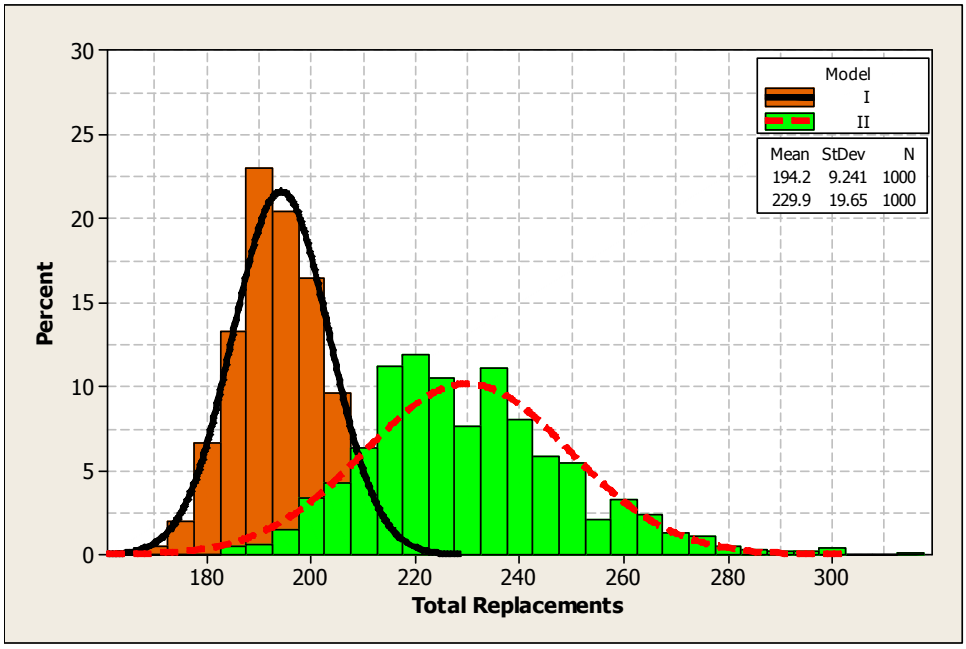


Figure 27: Sample distribution of estimated component replacements when the objective is a 5% reduction in predicted failures five years ahead for both failure prediction models.

A number of interesting observations can be made of Figure 27. The most obvious observation is the clear separation of the two distributions. The distribution of replacement actions for Model I is shifted towards fewer replacement components as compared to the distribution from Model II. Furthermore, the distribution of Model I is considerably narrower than that of Model II. Each distribution is also well represented by a normal distribution as shown fitted to each histogram in Figure 27. The parameters of these distributions are shown in Table 16.

Table 16: Summary of distribution parameters from Figure 27.

Distribution Parameter	Model I	Model II
Mean	194.2	229.9
Standard Deviation	9.24	19.65
Anderson-Darling Statistic	1.66	5.12
Anderson-Darling *p*-value	< 0.005	< 0.005

Table 16 shows the mean and standard deviation for Model I are 15.5% and 53.0% less, respectively, than those for Model II. Furthermore, the Anderson-Darling *p*-value indicates that with greater than 99.5% confidence, the data are normally distributed. The Anderson-Darling statistic measures the error between the empirical cumulative distribution function and the fitted distribution and can be used for comparison rather than as an absolute measure of the fit [36]. This statistic shows that the normal fit is better for Model I than for Model II. Still, both distributions are adequately modeled by this choice.

Figure 27 shows only one of the failure scenarios, the remaining two are shown more compactly using the box plot format as illustrated by Figure 28.

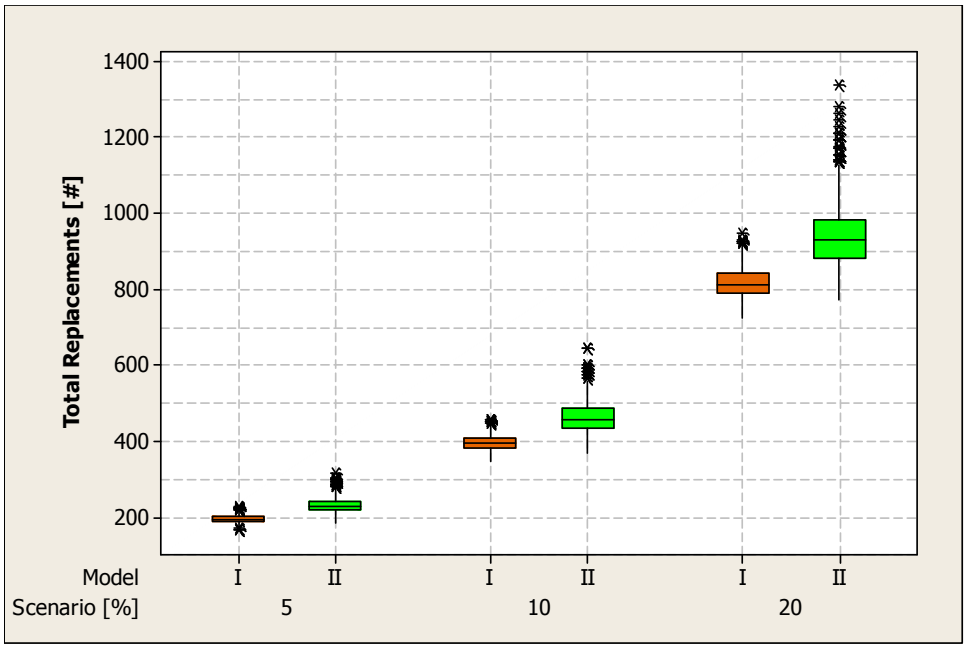


Figure 28: Distributions of estimated replacement rates for target failure reductions of 5%, 10%, and 20%, for both failure prediction models.

As Figure 28 shows, the distributions of replacements resulting from Model I are always displaced towards fewer replacements are compared to those of Model II. In addition, the Model I distributions are narrower for each failure reduction scenario. These points are more clearly seen by comparing the normal distribution parameters for each of the scenarios as shown in Figure 29.

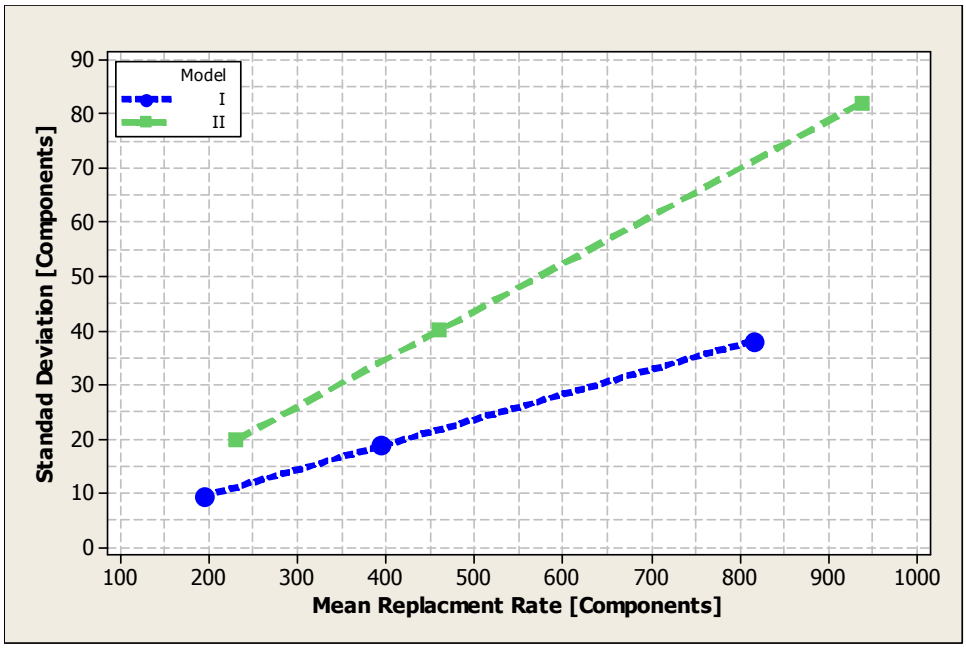


Figure 29: Fitted normal distribution parameters for distributions produced by estimated component replacement rates for each failure reduction scenario and failure prediction model.

Figure 29 clearly shows that as the mean replacement rate increases, so does the spread of the distribution as evidenced by the increase in standard deviation. This increase is not the same for the two models. In fact, the spread of the Model II distributions increases at a faster rate than those of Model I as shown by the difference in gradients of the fitted lines. In fact, Model I increases at only 52.4% of the rate of Model II. This is an interesting behavior as the randomness introduced by the Monte Carlo routine was virtually identical for the two models. However, as Figure 29 shows, the uncertainty at the output increases at a different rate in each model. Such evidence helps to show that there is benefit to having the additional information in the form of less uncertainty.

Figure 30, Figure 31, and Figure 32, show the upper confidence limits for each of the three failure reduction scenarios.

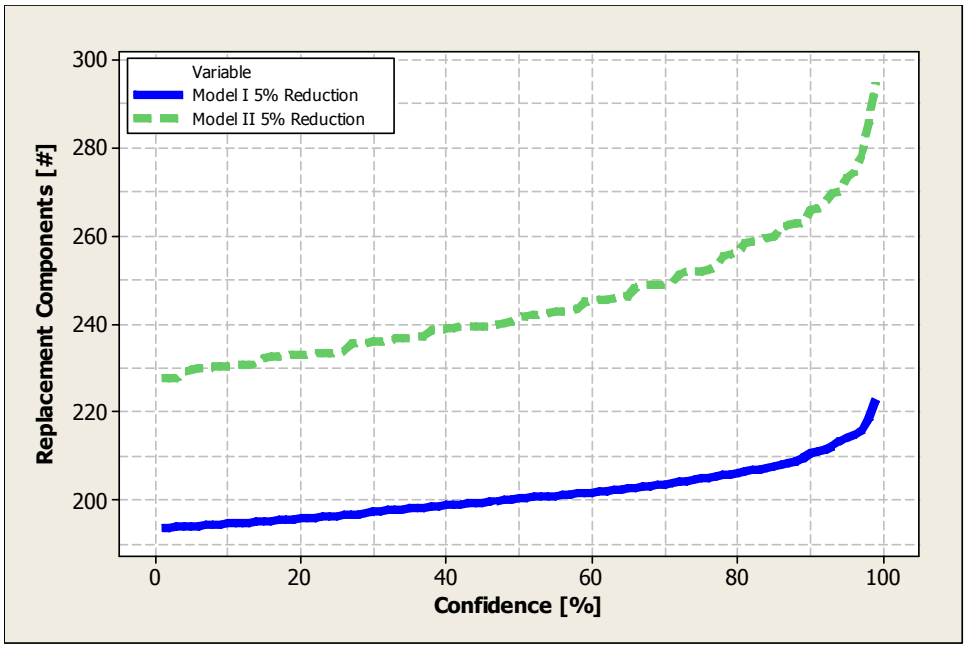


Figure 30: Estimated replacement actions that correspond to the upper confidence levels for Model I (—) and Model II (- - -) assuming a target reduction of 5%.

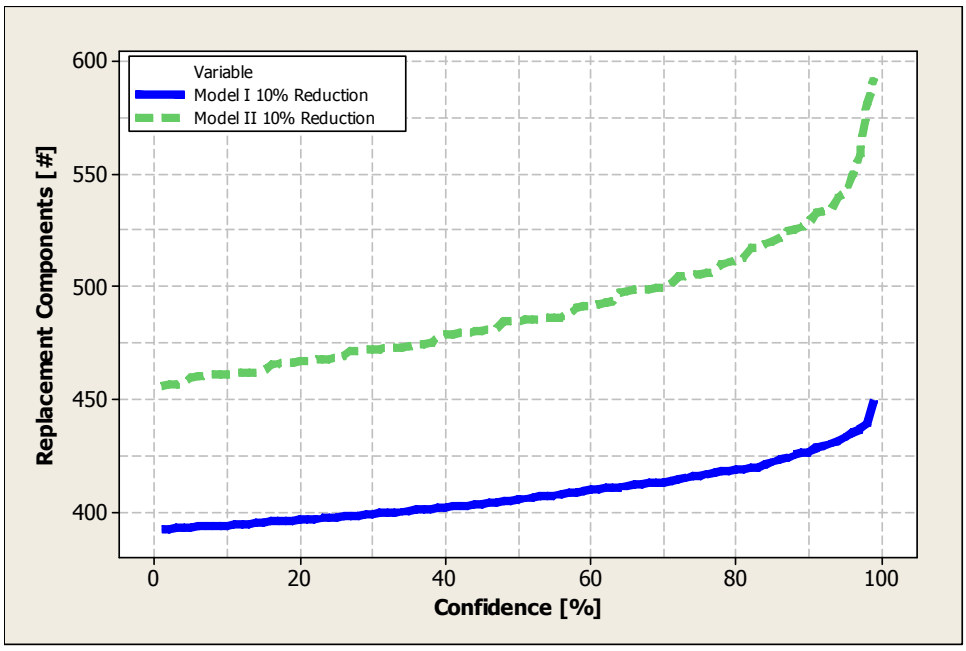


Figure 31: Estimated replacement actions that correspond to the upper confidence levels for Model I (—) and Model II (- - -) assuming a target reduction of 10%.

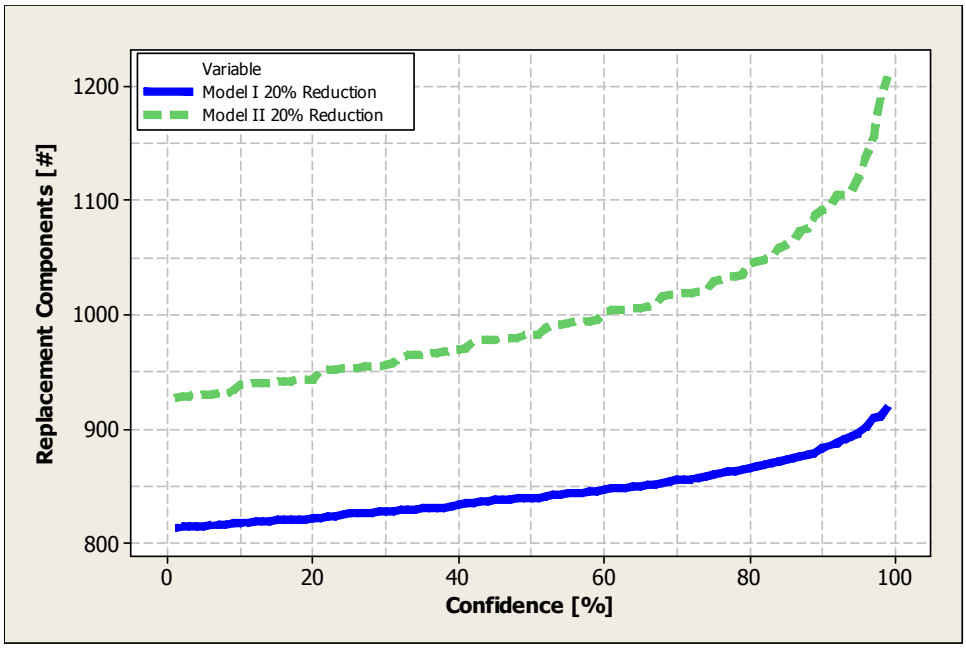


Figure 32: Estimated replacement actions that correspond to the upper confidence levels for Model I (—) and Model II (- - -) assuming a target reduction of 20%.

Figure 30, Figure 31, and Figure 32, each show that the upper confidence level of Model I is consistently less than that of Model II. In fact, Model I maintains substantially lower gradients than those produced by Model II. This is evidence that the efficiency of allocating the replacement components is higher with Model I. Also, the increasing gradients produced by Model II show that Model II must replace more vintage populations than Model I. Section 5.2.3 examines more closely this issue of efficiency.

5.2.3 Avoided Failures

Figure 33 shows the distributions for the expected number of avoided failures resulting from the replacement actions described in Section 5.2.2.

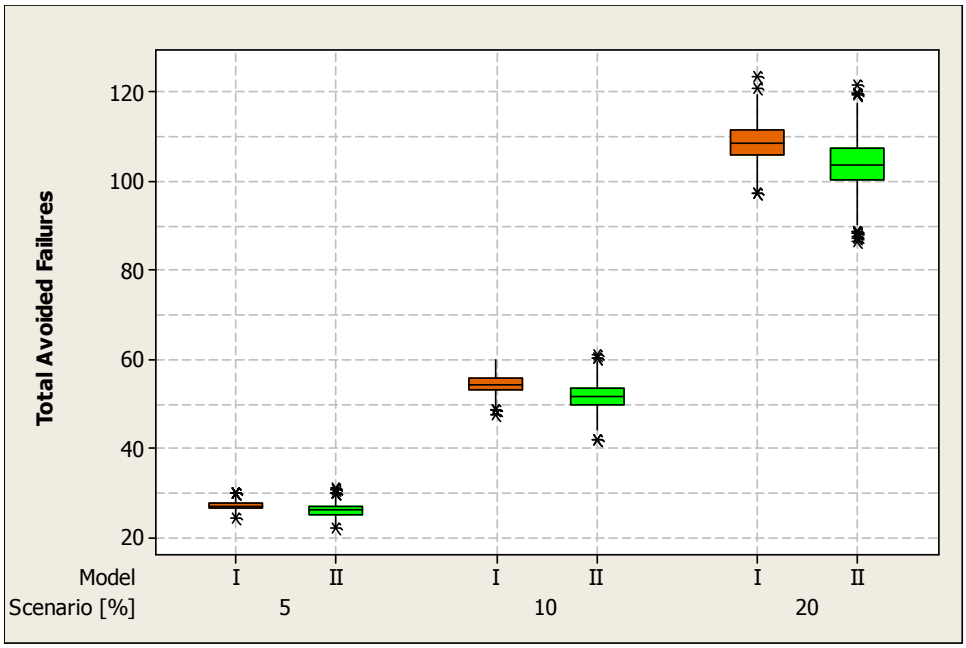


Figure 33: Distributions of avoided failures for target failure reductions five years ahead of 5%, 10%, and 20%, for both failure prediction models.

Figure 33 shows that in all cases, the predicted numbers of avoided failures is higher for Model I. This would indicate that failure predictions for Model I are higher than those of Model II. Yet, the previous section shows that Model I ultimately computes a lower replacement rate than Model II. This is the result of the higher replacement efficiency of Model I as shown in Figure 34. Clearly, the assumption used in Model II greatly impacts its ability to target replacement actions to the vintages that would yield the largest effect.

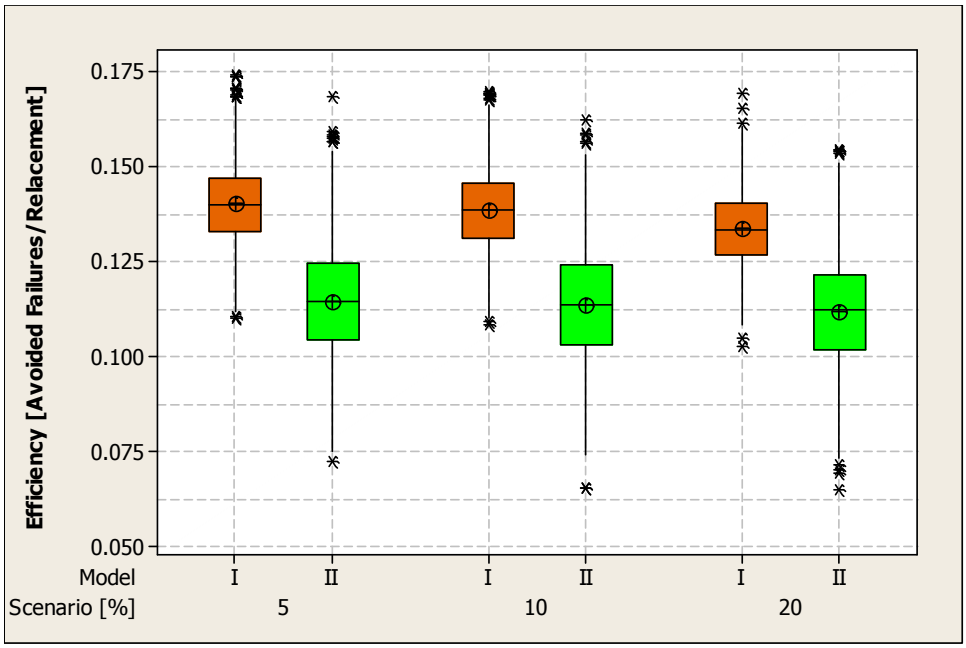


Figure 34: Distributions of replacement efficiency for target failure reductions of 5%, 10%, and 20%, for both failure prediction models.

5.2.4 Observations

Based on the results presented in the previous sections for Dataset 1, it is clear that for this dataset Model I allocates replacement components more efficiently than Model II. Furthermore, the predictions themselves are more narrowly distributed with Model I. The following section will explore the performance of these models on a dataset with different characteristics.

5.3 SYNTHESIZED DATASET 2

This section examines the results of Monte Carlo simulation for Dataset 2 (Appendix A) using both Model I and Model II. As in Section 5.2, the primary interest is the comparison of the results from the two models.

5.3.1 Simulation Setup

The Monte Carlo simulation parameters used for Dataset 2 are shown in Table 17. As in the case of Dataset 1, the randomization procedure employed in both models is identical and is based on the random sampling technique described in Section 4.2.2.

Table 17: Monte Carlo simulation parameters for each failure prediction model.

Parameter		Model I	Model II
Simulations [#]		1000	
Prediction Years		5	
Target Failure Reductions		5%, 10%, and 20%	
Dataset Generation Weibull Parameters	α	20.04	18.82
	β	2.01	1.90

5.3.2 Estimated Replacement Actions

Figure 35 shows an example of the distribution produced by each failure prediction model for the estimated component replacements needed to achieve a reduction in failures of 5%.

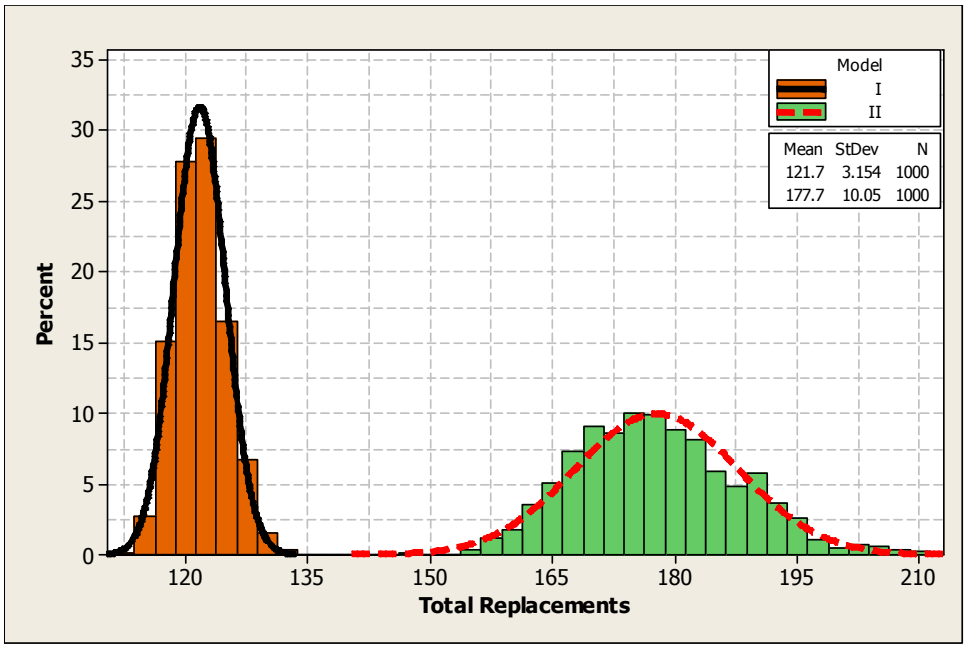


Figure 35: Sample distribution of estimated components replacements for a desired 5% reduction in predicted failures five years ahead for both failure prediction models.

Figure 35 shows a clear separation between the distributions of replacement actions produced by Model I and Model II. This separation is significantly larger than was observed using Dataset 1. In fact, the difference in the mean values is 56 components or approximately 31% of the estimated mean for Model II. On the other hand, the distribution from Model I is again quite narrow as compared to Model II's distribution. Comparing the estimated standard deviations shows that Model II's is 3.18 times larger than that of Model I. Unfortunately, comparing the estimated normal distribution parameters may not be as valid as it was for Dataset 1 (see Table 18).

Table 18: Summary of distribution parameters from Figure 35.

Distribution Parameter	Model I	Model II
Mean	121.7	177.7
Standard Deviation	3.15	10.05
Anderson-Darling Statistic	0.64	2.64
Anderson-Darling *p*-value	0.097	< 0.005

Table 18 shows that the results from Model I are not well represented by a normal distribution in this case. The lognormal distribution is an alternative to the normal distribution and the results of this fit are shown in Table 19.

Table 19: Summary of distribution parameters for lognormal distribution from Figure 35.

Distribution Parameter	Model I	Model II
Location	4.80	5.18
Scale	0.0259	0.056
Correlation	0.999	0.997
p-value	< 0.001	< 0.001

Based on the statistical parameters in Table 19, the lognormal distribution represents an improved fit for the data from Model I and still an excellent fit for the Model II data. The high correlation value translates to greater than 99.9% assurance that fit is not a random occurrence. In this distribution the location and scale parameters are the natural logs of the mean and standard deviation for the equivalent normal distribution.

The estimated component replacement rates for all scenarios are shown in box plot format in Figure 36.

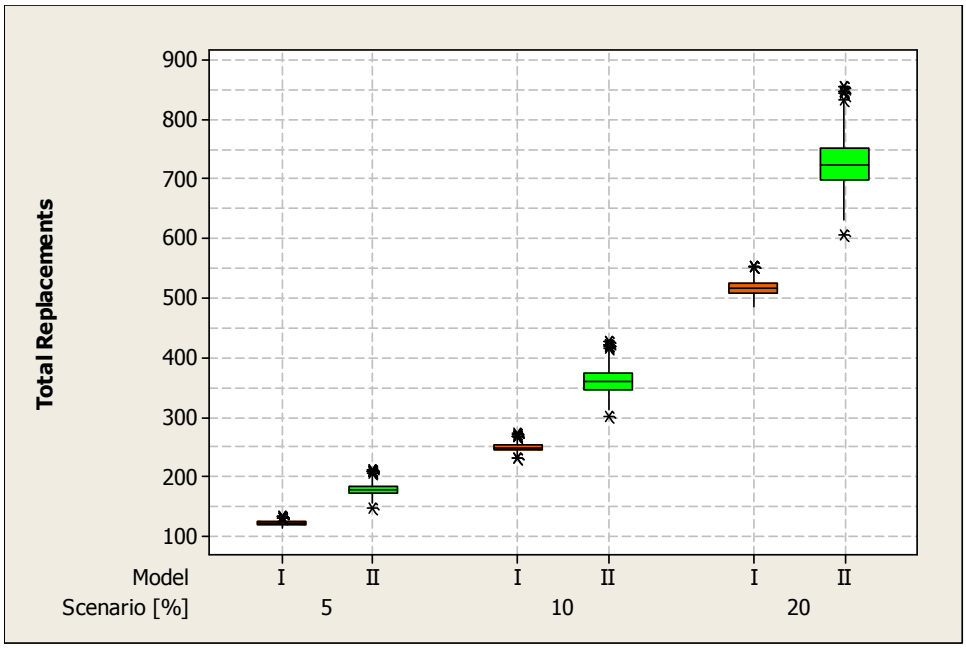


Figure 36: Distributions of estimated component replacement rates for target failure reductions five years ahead of 5%, 10%, and 20%, for both failure prediction models.

As Figure 36 shows, the distributions of replacement actions resulting from Model I are again displaced towards fewer replacements as compared to those of Model II. This difference is even more substantial than was found for Dataset 1. Also, the Model I distributions are again narrower for each failure reduction scenario. This can be seen graphically in Figure 37 as it shows the location and scale parameters of the lognormal distributions.

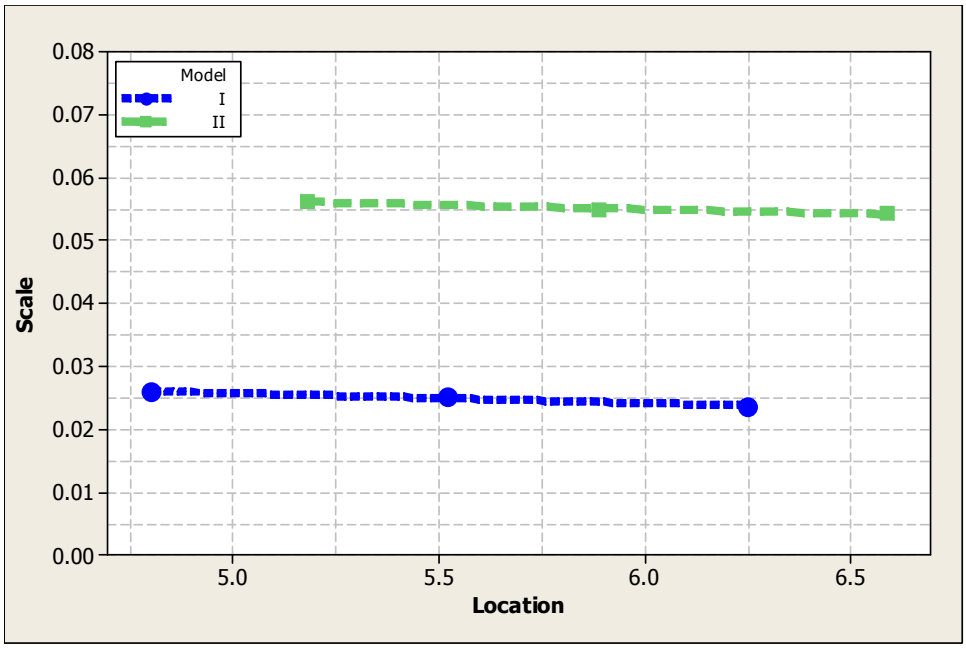


Figure 37: Fitted lognormal distribution parameters for distributions produced by estimated component replacement rates for each failure reduction scenario and failure prediction model.

Figure 37 shows a similar picture to that of Figure 29 in Section 5.2.2 except in this case the resulting slopes are considerably reduced. In fact, the slopes are both slightly negative. Still for a given location parameter the corresponding scale parameter for Model I is significantly less than the scale parameter for Model II. A closer inspection of data reveals that the scale parameter for both models is relatively independent of the location parameter, and more importantly, of the objective failure reduction. This behavior is different from that seen with Dataset 1.

On the other hand, this dataset behaves similarly to Dataset 1 in that the distributions for Model I correspond to significantly fewer replacement actions and are narrower. The differences in location and scale parameters are relatively constant between models averaging 0.361 and 0.030, respectively.

The replacements rates may also be examined from the perspective of upper confidence levels as shown in Figure 38, Figure 39, and Figure 40.

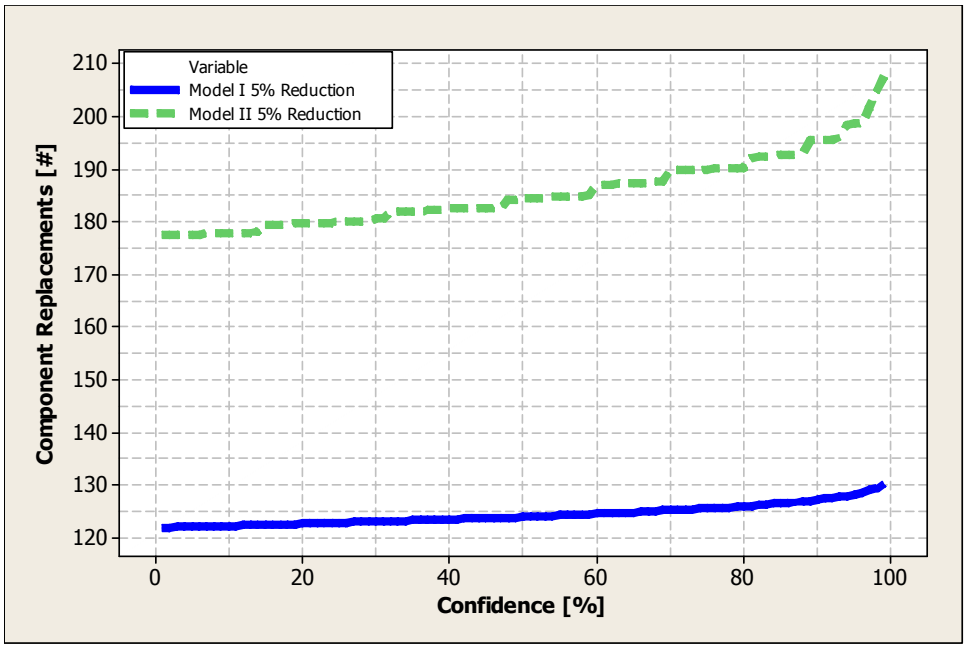


Figure 38: Estimated replacement actions that correspond to the upper confidence levels for Model I (—) and Model II (- - -) assuming a target reduction of 5%.

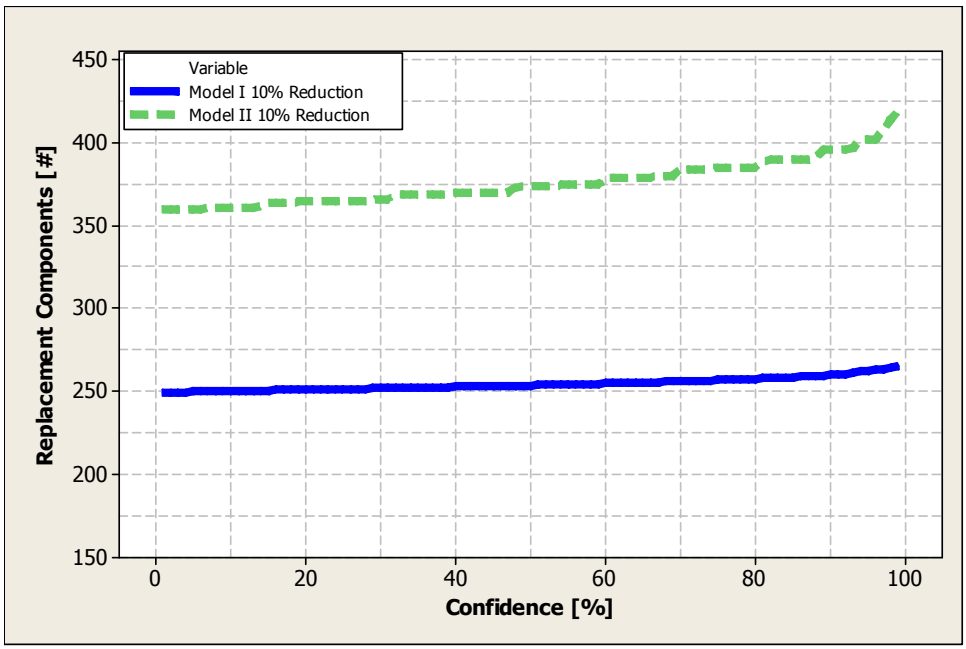


Figure 39: Estimated replacement actions that correspond to the upper confidence levels for Model I (—) and Model II (- - -) assuming a target reduction of 10%.

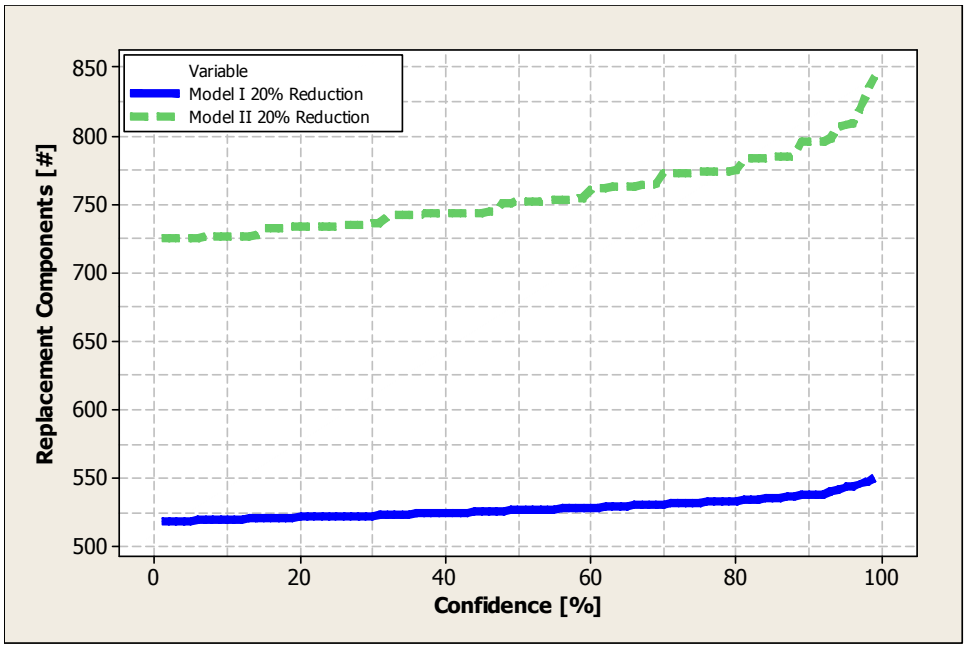


Figure 40: Estimated replacement actions that correspond to the upper confidence levels for Model I (—) and Model II (- - -) assuming a target reduction of 20%.

Figure 38, Figure 39, and Figure 40, each show that the upper confidence level of Model I is less than that of Model II as was seen in the case of Dataset 1. The differences in the gradients are even more pronounced than with Dataset 1. This is again evidence that the efficiency of allocating the replacement components is higher with Model I.

5.3.3 Avoided Failures

Figure 41 shows the distributions for the expected number of avoided failures resulting from the replacement actions discussed in Section 5.3.2.

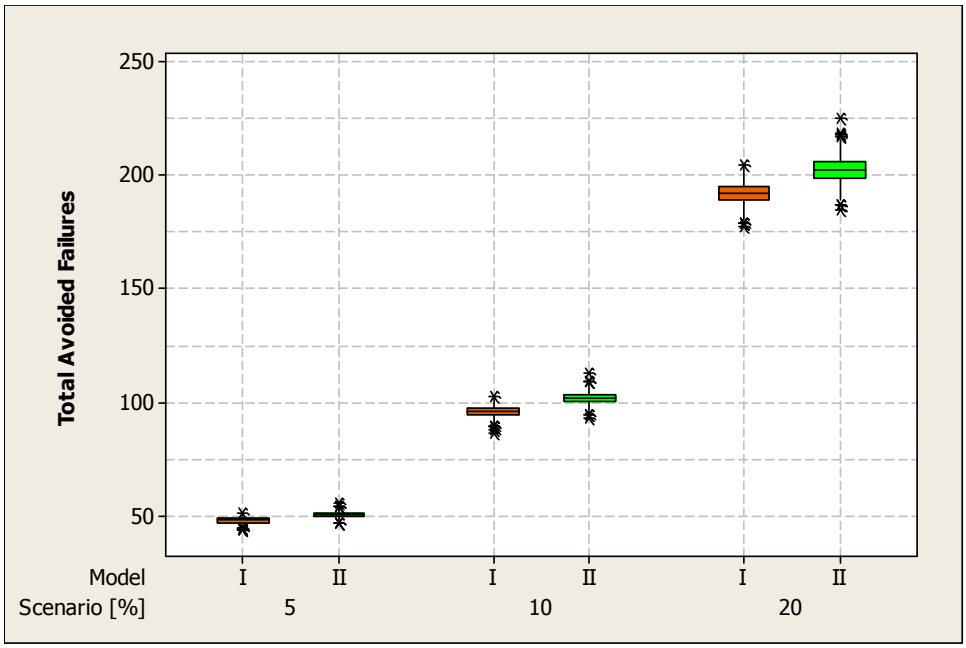


Figure 41: Distributions of avoided failures for target failure reductions of 5%, 10%, and 20%, for both failure prediction models.

Figure 41 shows a slightly different behavior as that of Figure 33 in Section 5.2.3 in that in all scenarios the failure avoidance rate for Model I is less than that of Model II. In addition, the distributions for all cases are quite narrow for both models. The differences in median values increase as the objective failure reduction increases.

Figure 42 shows the efficiency of allocating the replacement components is again significantly higher for Model I. Numerically speaking, Model I is approximately 38% more efficient than Model II.

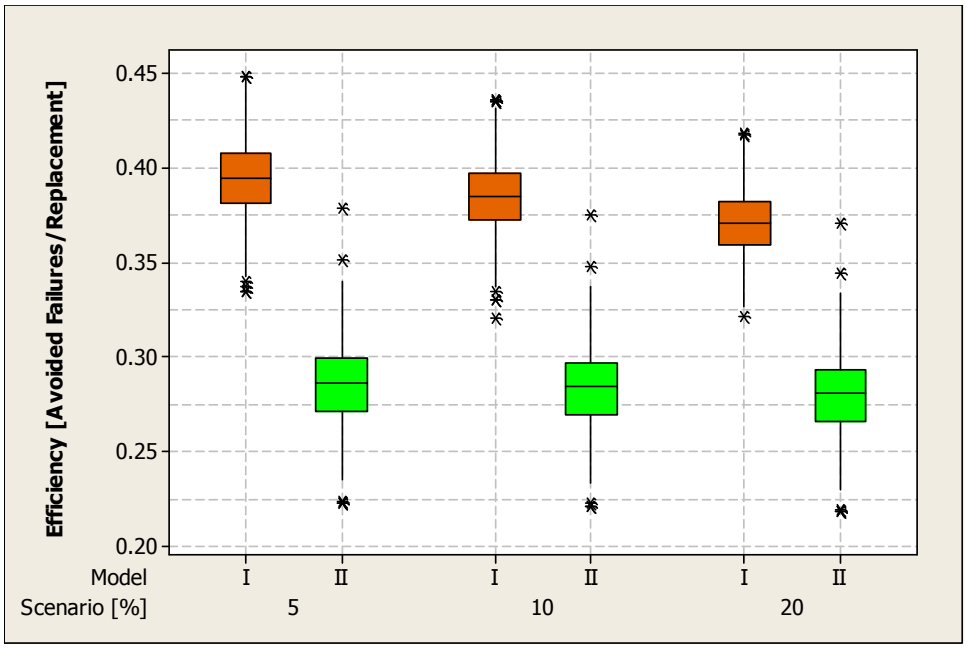


Figure 42: Distributions of replacement efficiency for target failure reductions of 5%, 10%, and 20%, for both failure prediction models.

5.3.4 Observations

As in the case of Dataset I, Model I again demonstrates that it is more efficient at allocating replacement components for each of the failure scenarios. Similarly, Model I produces narrower distributions than Model II. Still, it is clear from both datasets that in the absence of the age information required by Model I, Model II produces useful predictions that can be utilized by asset managers for planning purposes. Model II is simply more conservative in what it promises as compared to Model I. The following section utilizes only Model II to observe its behavior using the field data discussed in Section 3.3.

5.4 FIELD DATA

This section describes the results of Monte Carlo simulation using the field data presented in Section 3.3.

5.4.1 Simulation Setup

The Monte Carlo simulation parameters used for the field dataset are shown in Table 20. The randomization procedure employed in for this dataset is based on the multiple distribution random sampling technique described in Section 4.2.2.1.

Table 20: Monte Carlo simulation parameters for Model II.

Parameter		Model II
Simulations [#]		1000
Prediction Years		5
Target Failure Reductions		Variable (5%) and Fixed
Dataset Generation Weibull Parameters	α	Variable (Depends on annual failure data)
	β	2.15

An example of the resulting distribution of random inputs for one year is shown in Figure 43. Note that this distribution is unique for that particular year of data. Each year will have its own distribution that is adjusted to the initial failure estimates for that year.

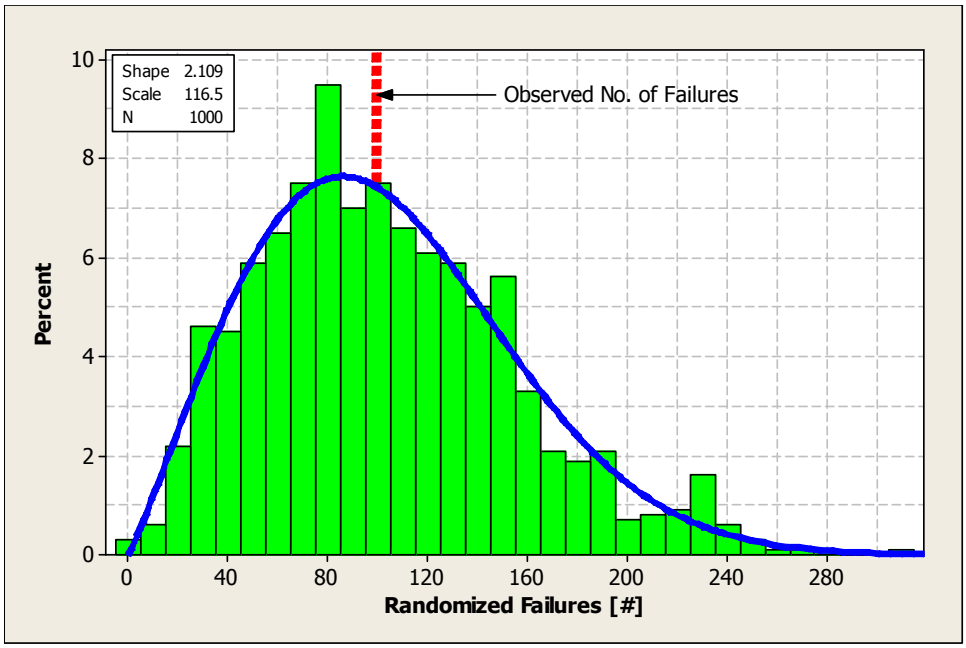


Figure 43: Histogram of randomized failure input for 1988 failures. The observed number of failures was 100 and the original estimated number of failures is 102.89.

Figure 44 shows the random distributions of inputs for all years of data using the box plot format described in Section 5.1.

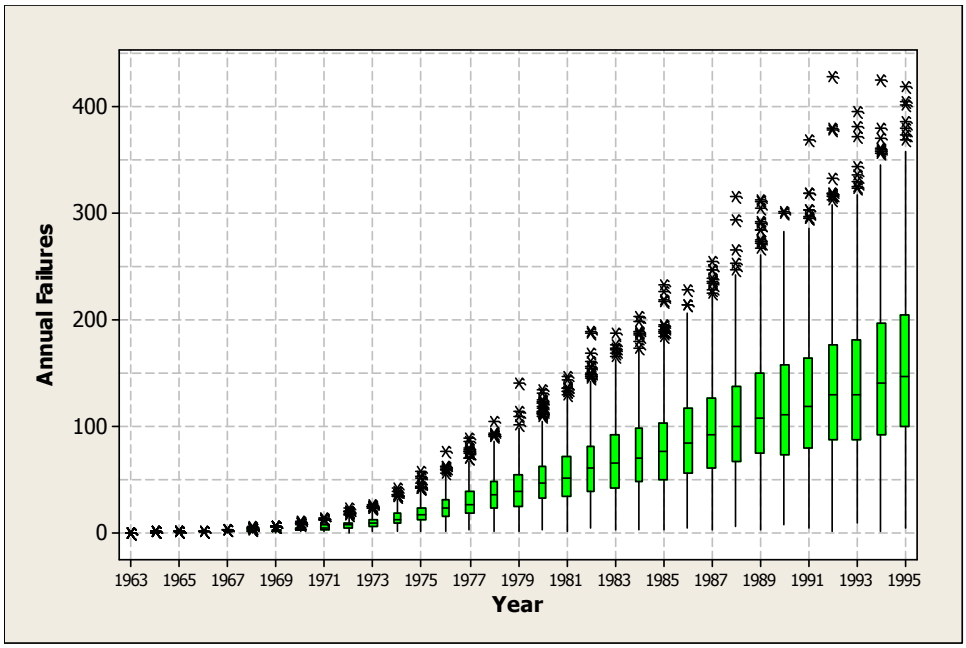


Figure 44: Box plot showing the distributions for each year's annual failures that will be used as input data for the Monte Carlo simulation.

It is apparent from Figure 44 that as the number of failures increases each year's corresponding distribution widens. This leads to the inclusion of a greater range of values in the Monte Carlo simulation for the later years of data.

5.4.2 Estimated Replacements

Using the data shown in Figure 44 as the input data to Model II of the failure prediction facet, it is possible to perform three tasks on each of the randomized datasets:

1. Predict failures for next year (one year past end of dataset) assuming no changes to population.
2. Compute the required number of components (miles of cable in this example) to replace to achieve the target number of failures for next year.
3. Predict failures for next year assuming computed component replacements are performed

These computations will be demonstrated using two different methods of selecting the target failure levels. The first is a fixed value equal to a 5% reduction in the original dataset's prediction for next year of 162.4 failures. This fixes the target failure level at 154.3 failures. In the second case, the failure reduction is again 5% but is based on each random dataset's own prediction for next year (as in Sections 5.2 and 5.3) and is, therefore, variable between different input datasets. These two scenarios represent two practical approaches that a utility could take in addressing the issue of how many components to replace.

Figure 45 shows the results of each computation in histogram form for the fixed desired failure level.

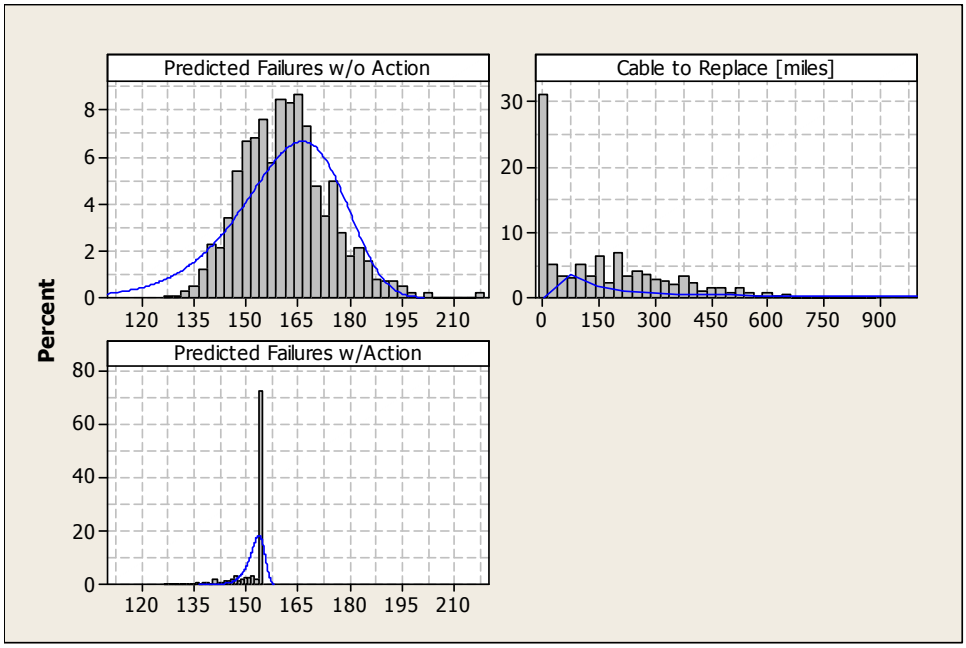


Figure 45: Histograms showing Monte Carlo simulation results for predicted failures without action (upper left), predicted failures with action to reduce failures to 154.3 (lower left), and required cable replacements (upper right).

In comparing the predicted failures with and without action, as shown in Figure 45, the two distributions are quite different. This is expected, given the objective of the

replacement actions is to adjust the failure level for all datasets to 154.3. In the *Predicted Failures w/Action* distribution, data points are located below the desired level since some of the datasets produced initial predictions that were already below the desired failure level. This is confirmed by the large percentage of zero replacement actions shown in the *Cable to Replace* distribution. This distribution also displays a long tail for the same reasoning. Some datasets produced initial predictions that were much higher than the target failure level and so required a large number of replacement components.

Figure 46 shows the same set of results for the case where the target failure level is variable and depends on the particular dataset.

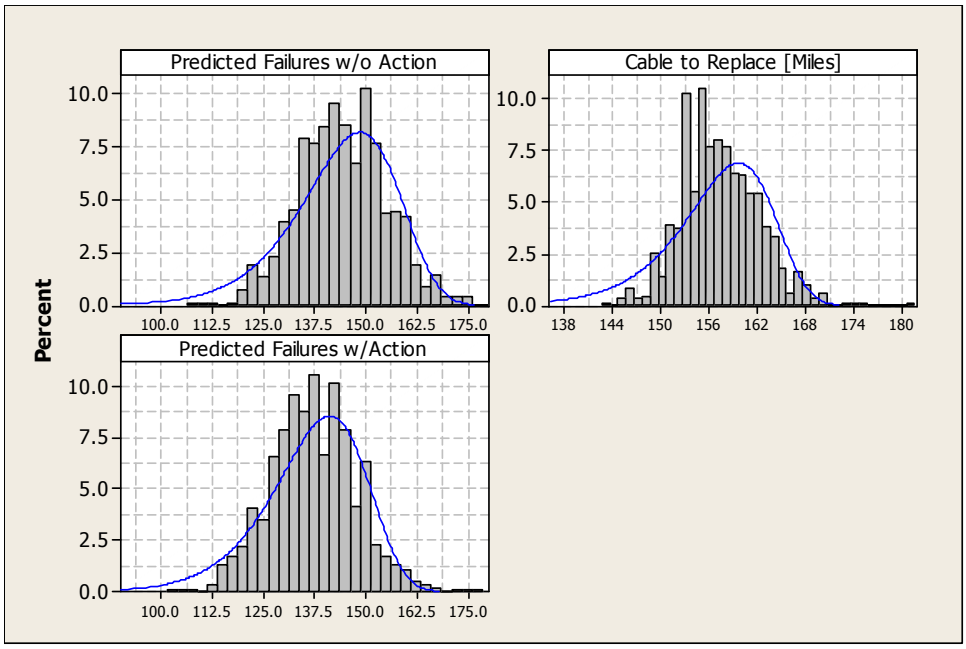


Figure 46: Histograms showing Monte Carlo simulation results for predicted failures without action (upper left), predicted failures with action to reduce failures by 5% (lower left), and required cable replacements (upper right).

In this case, all datasets will require replacement actions to meet their target failure reductions. In addition, the distributions for both sets of failure predictions are very similar, simply displaced by 5% from one another.

In comparing Figure 45 to Figure 46 the two distributions for *Predicted Failures w/o Action* are quite similar in shape. However, these distributions differ in terms of mean and variance as shown in Table 21. A much more striking difference can be seen in the *Predicted Failure w/Action* distributions. This difference is due to the different desired failure levels used in each simulation. Therefore, it is expected that the *Predicted Failures w/Action* for the fixed desired failure case would be a much tighter distribution than the variable case. In the latter case, the expectation is that that distribution would be nearly identical to the distribution of *Predicted Failures w/o Action*. In the last set of distributions (*Cable to Replace [miles]*) the distribution for the fixed failure case is weighted toward zero replacements. This arises from datasets where the initial failure prediction is already below the desired failure level. On the other hand, the variable case will always require some replacements as the goal is to always reduce the predicted failures by 5%.

Table 21: Summary of sample means and variances for the data shown in Figure 45 and Figure 46.

Target Failure Level	Output	Sample Mean	Sample Variance
Fixed	Failures w/o Action	161.4	159.9
	Failures w/Action	152.2	18.4
	Computed Replacements	163.7	28595
Variable	Failures w/o Action	144.5	119.5
	Failures w/Action	137.3	107.8
	Computed Replacements	157.3	23.4

The information for the variable case can be extended by considering the distribution of *Cable to Replace* in terms of the percentage of the total population, as shown in Figure 47. This figure shows that the median replacement cable rate is approximately 157 miles and this corresponds to approximately 4.2% of the total installed population.

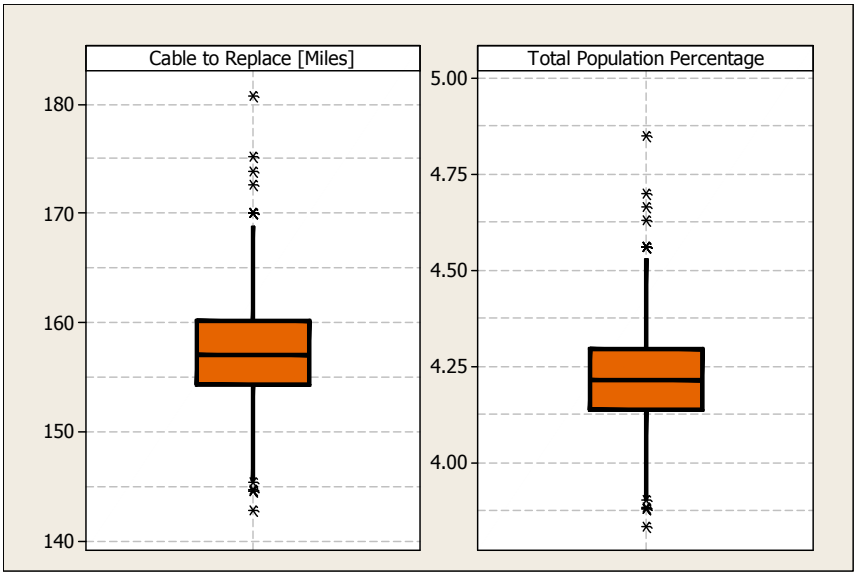


Figure 47: Amounts of cable to replace in 1996 in terms of miles (left) and percent of total cable population (right) to reduce the predicted failures by 5%.

5.4.3 Avoided Failures

As a measure then of the efficiency of replacement, Figure 48 shows the number of failures avoided by the cable replacements and the resulting failure reduction per mile replaced for the variable scenario.

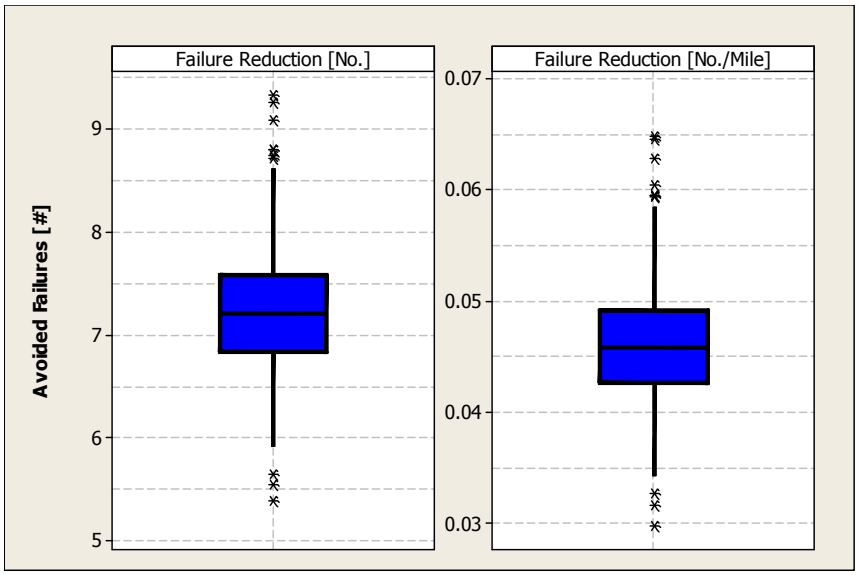


Figure 48: Failures avoided through cable replacement in number of failures (left) and failures per mile replaced (right).

Unfortunately for the system data used in this example, it takes more than 20 miles of cable replacements in order to avoid one failure in the next year.

Ultimately, it is possible to construct a replacement versus confidence plot. This allows a utility to identify the number of replacement actions needed to guarantee, with a desired confidence, that next year's failures do not exceed the target level. Examples of these plots are shown in Figure 49 and Figure 50, for the fixed and variable scenarios, respectively. Notice that in both cases that the number of replacements increases as the confidence increases. In theory, one could construct a curve for each target failure level of interest and then determine how achievable those cases are for different numbers of replacement actions. This allows one to manage risk at the desired level of tolerance.

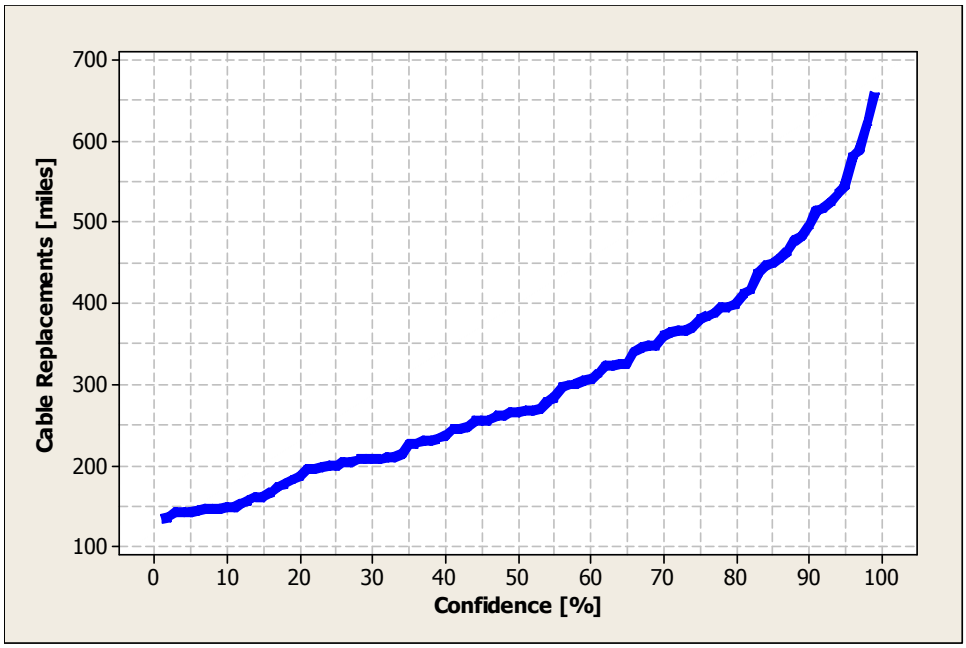


Figure 49: Cable replacements needed to guarantee the next year's failure performance meets or exceeds the target level with specified confidence for the fixed scenario.

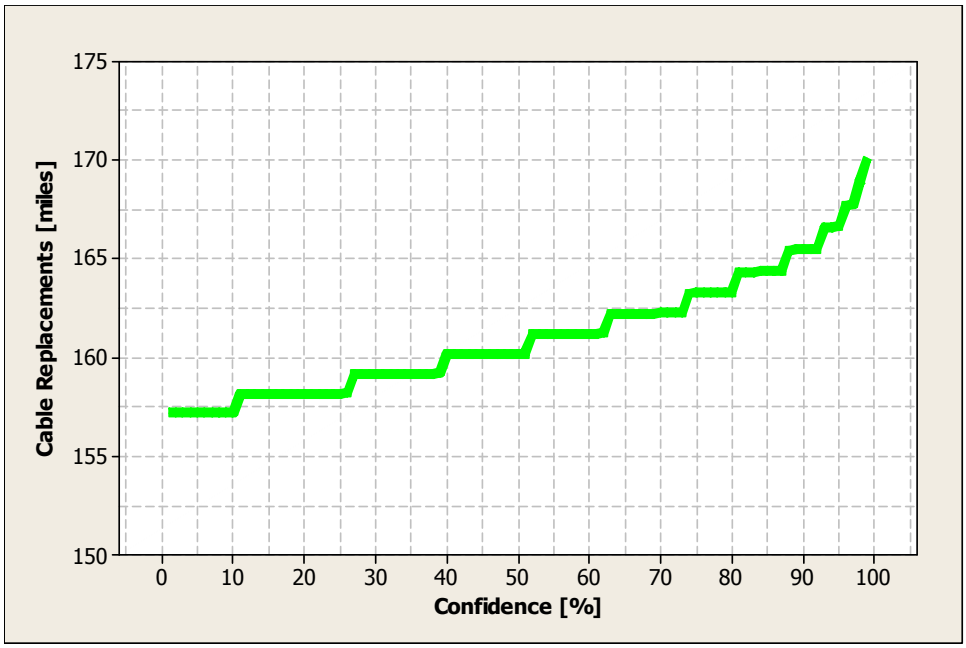


Figure 50: Cable replacements needed to guarantee the next year's failure performance meets or exceeds the target level with specified confidence for the variable scenario.

5.4.4 Observations

The study presented in this section demonstrates the flexibility of Monte Carlo techniques in handling different scenarios. The analysis of the results produced by Model II show that its allocation replacement actions is more efficient than randomly selecting cable from the overall population. This is a useful observation since the limited field data currently available to utilities will require the use of Model II until more complete data can be put together for use in Model I.

5.5 SUMMARY

This chapter presents results of Monte Carlo simulation for two synthesized datasets and one field dataset for an underground cable system. The comparison of results obtained from both failure prediction models shows that Model I produces narrower distributions for all output variables as well as lower estimates for replacement components. This again implies that Model I is able to allocate the replacement components such that a higher return in terms of avoided failures is achieved as compared to Model II. This is shown in the component replacements versus confidence plots. In the absence of the data needed for Model I, Model II still provides reasonable predictions and estimates that are simply less optimal than those of Model I.

PART I SUMMARY

Part I describes the research performed in connection with the failure prediction facet. This facet includes two Weibull-based failure prediction models that are each able to predict future fail trends for a homogeneous population of components. To do so, these models rely solely on historical data that would include annual numbers of installations, replacements, and failures, either for the whole population or separated by vintage.

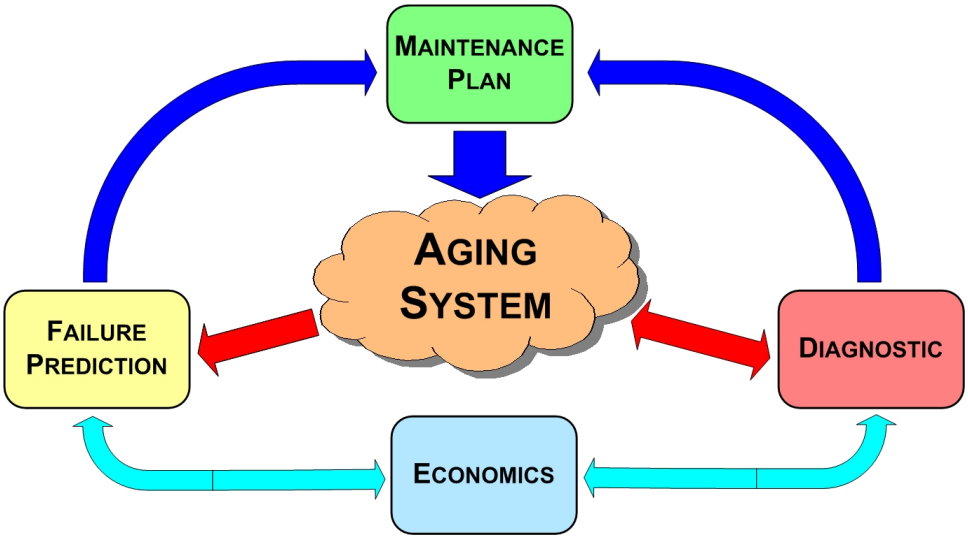
Several example simulations were performed using both synthesized and field data. The synthesized data allowed for the comparison of the two models so that the benefit of the additional information in Model I could be determined. As a result of these simulations, it was shown the Model I provides greater efficiency in the allocation of replacement actions in that each replaced component leads to a greater reduction in predicted failures. Model II was shown to be approximately 30% less efficient. While a rather significant difference, this does not negate the usefulness of Model II in situations where Model I cannot be used. These comparisons show that minor investment in failure tracking and record keeping would produce substantial savings in replacement strategies based on failure prediction. Until enough time passes for such data to be gathered in large enough quantities, Model II can be used with some reduced efficiency.

Simulations were also undertaken using actual field data obtained for an underground cable system. The results of this investigation showed that the choice of the Weibull distribution is consistent with the real behavior of power system equipment. This is vital information as the wrong choice of distribution would lead to highly inaccurate estimates and subsequent predictions.

The usefulness of the stochastic simulation technique will be made even more apparent in Part II and Part III since these each deal with probabilistic situations. In fact,

as shall be discussed in Part II, the use of diagnostics adds significantly to the uncertainty included in implementing a reliability enhancement program.

PART II: DIAGNOSTIC FACET



An accurate condition assessment of a component via a diagnostic test procedure requires a complete understanding of all the processes that can produce aging and, ultimately, failure in that device. For most components, this is an impossible feat to accomplish and, as a result, the available diagnostic tests do not always provide a perfect assessment of the component's current state. Even though these imperfections exist, diagnostics are still useful tools that can provide additional guidance in developing maintenance plans, provided their limitations are understood. The three questions that arise are:

(1) How to evaluate diagnostic tests?

(2) How to determine the diagnostic accuracy?

(3) What is the effect of the diagnostic program on the reliability of the system?

The diagnostic facet addresses each of the above questions and provides the feedback needed to determine whether or not a diagnostic test should be employed. The first question requires suitable analysis techniques in order to correlate diagnostic measurements with actual component service performance. Ideally, the diagnostic test should be capable of distinguishing components that will and will not fail within a reasonable time horizon. One of the goals of the NEETRAC CDFI project is to establish the effectiveness of diagnostics for underground cable systems. As part of this project, several techniques have been developed or adapted to evaluate the commercially available diagnostic techniques currently used for assessing underground cable. These analysis techniques are discussed in Chapter 7 and include: (1) performance ranking, (2) diagnostic outcome mapping, (3) Weibull analysis, (4) survivor analysis, and (5) classification.

However, prior to the discussion on the techniques themselves it is necessary to examine the meaning and interpretation of "accuracy." Specifically, its meaning and how

one may compute it. This will form the basis of the answer to the second question mentioned above.

CHAPTER 6: DIAGNOSTIC ACCURACY

Diagnostic accuracy is a vital component to the process of developing a maintenance plan that includes decision making that is the result of diagnostic testing. The definition of accuracy in diagnostic testing is the percentage of tests whose resulting assessments correctly describe the true condition of the components tested. Unfortunately, this process is not straightforward and is the subject of this chapter.

In the case of underground cables, the primary interest is in determining which cable segments will and will not fail within a given time horizon. The latter case corresponds to the "good" components while the former case to the "bad" components. These "bad" components are expected to be more susceptible to failing as compared to the "good" components. Still, susceptibility does not guarantee that a failure will occur within an arbitrary time horizon. In any failure process there is a strong dependence on time. The components that are diagnosed as "bad" do not fail at the same time nor do the "good" components remain failure free indefinitely. This implies that the point in time at which the accuracy is computed will greatly affect the result.

There are two types of accuracies that are of interest:

- Overall Accuracy – For each test performed, what percentage of tests yielded results that correctly matched the component's condition? If the component was "bad," did it fail? If the component was "good," did it not fail?
- Condition-Specific Accuracy – For each possible component condition, what percentage of them were correctly diagnosed as having that condition?

The above accuracies are subtly different in their definitions but tremendously different in their implications. The primary difference between the two is how the group of tested components is subdivided. The first type of accuracy looks at all the

components together. This accuracy is best used as a means of comparing tow different diagnostics. The second type, condition-specific accuracy, examines the accuracies within the smaller groups (i.e., how many components diagnosed as "good" are actually in good condition and how many bad (degraded) components were mistakenly diagnosed as "good"). This accuracy is needed when the economic benefits are examined since the economic consequences of the wrong diagnosis of a "good" component are very different from those resulting from the misdiagnosis of a "bad" component. Fortunately, the two accuracies are mathematically related.

Consider the following example: suppose in a test of 100 components it was known before the test that 80 of them were in good condition and the remaining 20 were degraded. After testing the entire population the results in Table 22 were obtained.

Table 22: Summary of diagnostic testing results for the example of 100 components.

True Condition	No. of Components	No. Diagnosed as "Good"	No. Diagnosed as "Bad"
Good	80	64	16
Bad	20	2	18

Table 23 shows the resulting overall and condition-specific accuracies computed for the data shown in Table 22.

Table 23: Overall and Condition-Specific accuracies for "good" and "bad" components.

True Condition	Overall Accuracy [%]	Condition-Specific Accuracy [%]
Good	82%	80%
Bad		90%

There are a number of observations that can be made about Table 22 and Table 23. First, the "good" and "bad" groups are different fractions of the whole population. Second, the two groups have different condition-specific accuracies. The resulting overall accuracy is then a weighted average of the two Condition-Specific accuracies. This weighting is determined by the relative sizes of the two groups. One may also determine the condition-specific accuracies from the overall diagnostic accuracy and relative sizes of the two groups. These accuracies have tremendous importance for the Economics facet as discussed in Part III. However, the phenomenon itself is explained using Bayes' Theorem as described in the following section.

6.1 BAYES' THEOREM AND CONDITIONAL PROBABILITY

The application of Bayes' Theorem and conditional probability can be shown with an example: Consider a diagnostic test (T) able to provide two levels of condition assessment ("good" and "bad") and probabilities are associated with the performance of this particular diagnostic. If the diagnostic test T is perfectly accurate, then the probability of a positive test outcome (T > 0) on a "good" (G) component should be one. Similarly, the probability of the negative test outcome (T < 0) on a "bad" (B) component:

$$P(T>0|G)=1 \qquad P(T<0|G)=0, \qquad P(T<0|B)=1 \qquad P(T>0|B)=0. \tag{6.1}$$

Bayes' theorem relates conditional probabilities for stochastic events X and Y [37],

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}. \tag{6.2}$$

Using Bayes' theorem, the probability of diagnosing a "good" (G) component would indeed be the result of the positive test can be expressed as a function of the test accuracy:

$$P(G|T>0) = \frac{P(T>0|G)P(G)}{P(T>0)} = \frac{P(T>0|G)P(G)}{P(T>0|G)P(G)+P(T>0|B)P(B)}. \tag{6.3}$$

Probabilities *P(G)* and *P(B)* represent the probabilities of randomly selecting a "good" (G) or "bad" (B) component from the population and so *P(G)* + *P(B)* = 1. It is, therefore, clear that the probability of correct diagnosis will depend on the diagnostic accuracy as well as the concentration of "good" (G) and "bad" (B) components in the population. Considering a perfectly accurate diagnostic test, then the probability that a positive test would correctly identify a "good" component is 100%:

$$P(G|T>0) = \frac{P(T>0|G)P(G)}{P(T>0|G)P(G)+P(T>0|B)P(B)} = \frac{1\cdot P(G)}{1\cdot P(G)+0\cdot P(B)} = \frac{P(G)}{P(G)} = 1. \tag{6.4}$$

Similarly, the probability that a negative test would correctly identify a "bad" component is 0%.

If, however, the assumption is made that the diagnostic test is only 50% accurate (no useful diagnostic information is extracted from the tests), then the probability that a

negative test would correctly identify a "good" component would be reduced to the probability of randomly selecting a "good" sample from the component population:

$$P(G|T>0)=\frac{P(T>0|G)P(G)}{P(T>0|G)P(G)+P(T>0|B)P(B)}= \tag{3.5}$$

$$=\frac{0.5\cdot P(G)}{0.5\cdot P(G)+0.5\cdot P(B)}=\frac{0.5\cdot P(G)}{0.5}=P(G).$$

Figure 51 shows the dependence of the probability of correctly identifying a "good" component on the probability of randomly selecting a "bad" component from the total population. This is equivalent to plotting the condition-specific accuracy of the "good" as a function of the size of the "bad" portion of the population and overall diagnostic accuracy.

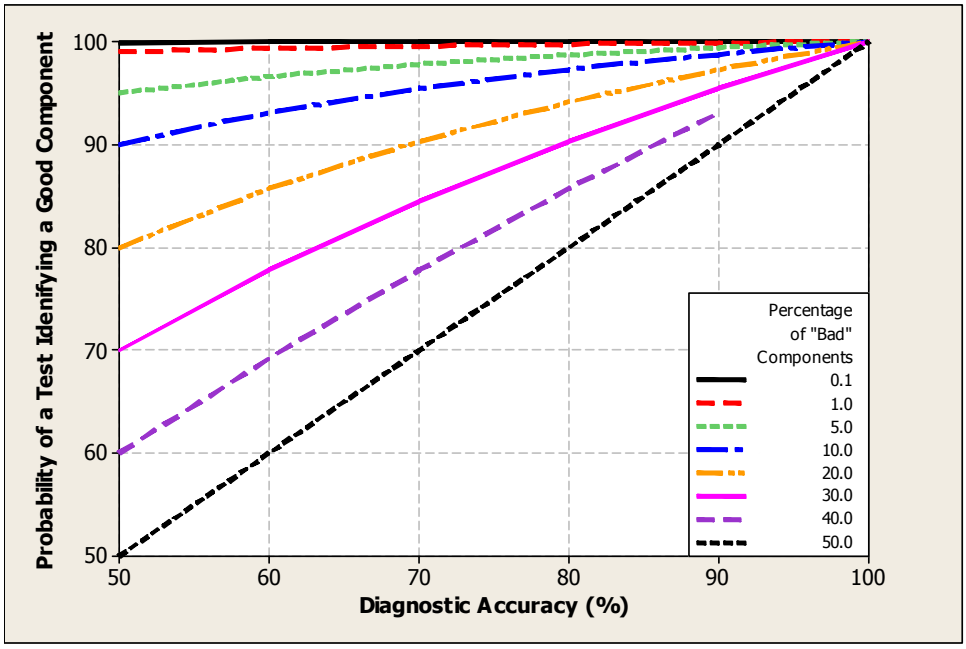


Figure 51: The probability that a positive test would correctly identify a "good" component is shown here as a function of diagnostic accuracy and proportion of the "bad" components in the population.

According to Figure 51, as the number of "bad" components decreases, the condition-specific accuracy increases for the "good" components. This is because there are so few "bad" components that diagnosing all of them as "good" leads to a high condition specific accuracy. On the other hand, when the population is 50% "bad" and 50% "good" then the condition-specific accuracies are both equal to the overall diagnostic accuracy.

Evaluating and interpreting the accuracies of different diagnostic tests is important in determining whether or not they can be employed. This will be shown in Part III during the discussion of the economics facet.

6.2 WHY ARE DIAGNOSTICS NOT "PERFECT"

It has become clear through the CDFI project for the case of underground cables that the interpretation of diagnostic data is a complicated process. Diagnostics encounter difficulty when trying to classify components whose conditions lie somewhere between "very bad" and "very good." The problem is that diagnostics generally only measure certain characteristics of, generally, one of the known failure mechanisms. These characteristics only provide an indication of the condition of the cable with respect to that particular failure mechanism. Other failure mechanisms are, in general, also present and interact with the particular one the diagnostic attempts to assess. This fact, in addition to others, causes significant overlap to occur between the distributions of measurements for components that fail and those that do not. With such overlap it can be impossible to decide whether the component belongs in the "good" or "bad" class. Figure 52 shows an example in which the boundary between the distributions is unclear because of overlap.

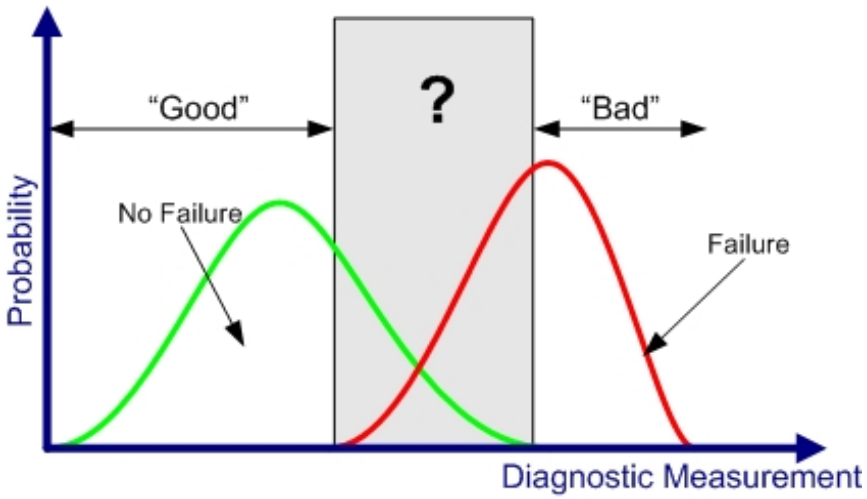


Figure 52: Sample distributions for diagnostic measurements where there is significant overlap between measurements from components that fail and those that do not.

As a result of the ambiguity shown in Figure 52, it is virtually impossible in practice for a diagnostic to achieve an overall accuracy of 100%. Yet diagnostics in other fields (including medicine) that are less than 100% accurate have proven to be useful despite their imperfections. The key is to know what the accuracy is so that decisions can be made that give proper consideration to the risks involved.

As mentioned above, evaluating diagnostic accuracy is a challenging task. The following sections describe the various techniques that have been developed for this purpose.

6.3 METHODS FOR COMPUTING DIAGNOSTIC ACCURACIES

Two methods can be used for the calculation of overall diagnostic accuracies depending on the amount of information available for the analysis:

1. "Bad Means Failure" approach (Method 1).
2. Probabilistic approach (Method 2).

The later method requires additional information on the service performance of components diagnosed in the past using the same diagnostic technique. This information may or may not be available and so in its absence Method 1 may be used. Each method is discussed in the following sections.

6.3.1 Method 1 – "Bad Means Failure" Approach

The "Bad Means Failure" approach relies on the assumption that all components diagnosed as "bad" must experience a failure between testing and the time of analysis. Likewise, the "good" components cannot experience failures during the same period. For example, if the analysis is conducted two years after testing, then the assumption must be made that all the "bad" components should have failed by that time while all the "good" components should still be in operation.

The procedure to compute the overall diagnostic accuracy is as follows:

1. Determine the number components that belong to each of the following groups:
    a. G_g – Components diagnosed as "good" and did not fail in service.
    b. G_b – Components diagnosed as "good" and did fail in service.
    c. B_g – Components diagnosed as "bad" and did not fail in service.
    d. B_b – Components diagnosed as "bad" and did fail in service.
2. Compute the overall diagnostic accuracy using:

$$G = G_g + G_b,$$

$$B = B_g + B_b, \tag{6.6}$$

$$\text{Overall Diagnostic Accuracy} = \frac{G_g + B_b}{B + G} \times 100\%.$$

The above procedure is somewhat simpler than the method that is described in the following section.

6.3.2 Method 2 – Probabilistic Approach

The assumption that all components diagnosed as "bad" failed by the time of analysis and all the components that were "good" remained so until the analysis generally leads to poor diagnostic accuracies. This is especially true for components diagnosed as "bad" since they do not all fail immediately after they are returned to service. In reality, there is a probability of failure associated with each "bad" component that is a function of time. As time passes, the probability of failure for each "bad" component increases. The same applies to components diagnosed as "good" but their probabilities are substantially lower than those components diagnosed as "bad". Depending on the time of analysis, be it one year, two years, or more, after testing, the expected number of components that would fail will be different. Equally, there are an expected number of components that should not fail within the chosen time frame. This expectation is based on the number of "good" and "bad" components and represents the basis for the probabilistic approach for computing diagnostic accuracies.

In order to use this probabilistic approach additional information on the service performance of components previously tested using the same diagnostic technique is needed. This could come from a diagnostic provider's database of measurements or follow-up on service performance from an earlier pilot program conducted by the utility. Using this data, it is possible to construct probability curves that represent the percentage of components that should fail given a particular diagnostic classification and time frame. Unfortunately, these curves cannot identify which components within a particular diagnostic class will fail but they can provide an idea of what should be expected. This concept is very useful as it represents a prediction of the future service performance of the tested population. An example will be used to illustrate this further.

The data needed to construct the probability curves for an underground cable diagnostic that uses a five level classification scheme (1-5 with five representing the

worst condition and highest probability of failure) is available. Condition classes one and two imply zero probability of failure within 10 years. As part of this diagnostic provider's normal reporting process the data needed to construct the probability curves for condition classes 3-5 is included. Using a parametric distribution (in this case the Weibull distribution) the resulting probability curves for this data are shown in Figure 53.

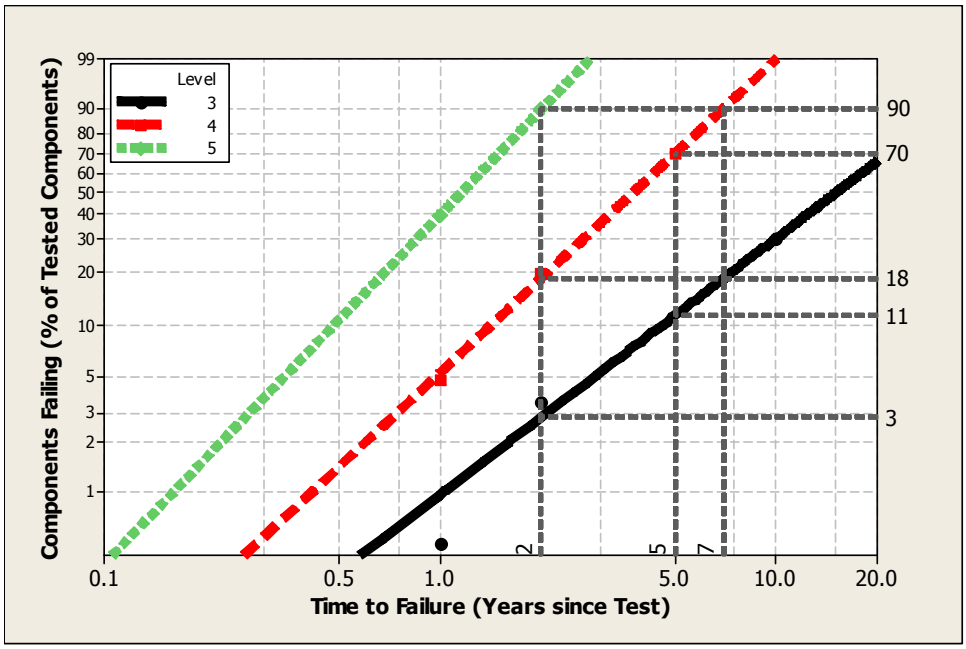


Figure 53: Weibull curves of time to failure for conditions three, four, and five. It is assumed that components with conditions one or two should not fail.

Figure 53 may then be used to predict the percentage of components from each condition class that should be expected to fail within a given time horizon. For example, within two years of testing approximately 3% of class threes, 18% of class fours, and 90% of class fives should fail. Table 24 shows some hypothetical data that demonstrates how Figure 53 can be used to predict the number of future failures.

Table 24: Example data from failure prediction.

Class	Components [#]	Failure Prediction After		
		2 Years	5 Years	7 Years
3	100	3	11	18
4	10	2	7	9
5	5	5	5	5
TOTAL	**115**	**10**	**23**	**32**

Using this data, it is possible to construct a probability curve for the total failures as shown in Figure 54.

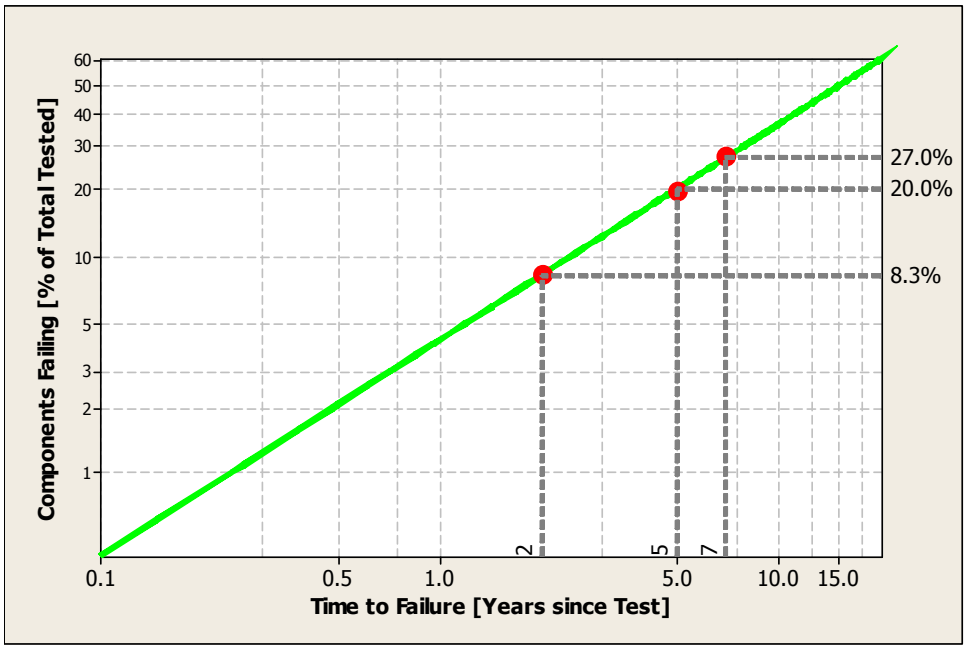


Figure 54: Total expected failures after testing for the hypothetical data in Table 24.

According to Figure 54, after two years, approximately 8% of the components would be expected to fail while 37% are expected to fail within 10 years. Comparing these predictions to the service performance of the tested population allows one to extract the diagnostic accuracies.

These accuracies are computed by combining the different classes that would correspond to a "bad" condition. In the example above, condition classes 3-5 belong in this category. The procedure for computing the diagnostic accuracy requires four steps:

1. Determine the number components that belong to each of the following groups as described in Section 6.3.1:
   a. G_g – Components diagnosed as "good" and did not fail in service.
   b. G_b – Components diagnosed as "good" and did fail in service.
   c. B_g – Components diagnosed as "bad" and did not fail in service.
   d. B_b – Components diagnosed as "bad" and did fail in service.
2. From the probability curves, determine the percentages of components diagnosed as "good" ($G = G_g + G_b$) and components diagnosed as "bad" ($B = B_g + B_b$) that were expected to fail during the given time frame. Denote these percentages as P_g and P_b, respectively.
3. Compute the expected number of failures for both "good" and "bad" components as shown in (6.7).

$$F_g = P_g G$$
$$F_b = P_b B. \tag{6.7}$$

4. Compute the overall diagnostic accuracy using (6.8) and (6.9)

$$N_g = \begin{Bmatrix} F_g + G_g & G_b \geq F_g \\ 2G_b + G_g - F_g & \text{otherwise} \end{Bmatrix},$$
$$N_b = \begin{Bmatrix} F_b + B_g & B_b \geq F_b \\ 2B_b + B_g - F_b & \text{otherwise} \end{Bmatrix}. \tag{6.8}$$

$$\text{Overall Diagnostic Accuracy} = \frac{N_g + N_b}{B + G} \times 100\%. \tag{6.9}$$

Both methods are illustrated with an example in the following section.

6.3.3 Diagnostic Accuracy Calculation Example

Consider as an example the underground cable diagnostic described in 6.3.2. This diagnostic was applied to a utility cable system seven years ago with the summarized findings in Table 25. Note that the data has been provided in the format shown in Table 25.

Table 25: Summary of diagnostic test findings.

Condition Class	**Components** [#]	**Predicted Failures** [#]	**Actual Failures** [#]
1-2 "Good"	1290	0	71
3-5 "Bad"	349	168	99

6.3.3.1 Method 1

Using Method 1 as described in Section 6.3.1, the overall diagnostic accuracy is computed as,

$$\text{Overall Diagnostic Accuracy} = 100\% \times \frac{1219+99}{1639}$$
$$= 80\%.$$

6.3.3.2 Method 2

Using Method 2 as described in Section 6.3.2, the overall diagnostic accuracy for the same dataset using the probability curves shown in Figure 53 is:

$$\text{Overall Diagnostic Accuracy} = 100\% \times \frac{1219+99+181}{1639}$$
$$= 91\%.$$

The last term in the numerator comes from the fact that 181 components from the "bad" classes that should not have failed according to Figure 53.

In this example, Method 2 provides for a higher overall diagnostic accuracy that is more representative of what physically happens in service.

6.4 ACCURACY OF THREE OR MORE DIAGNOSES

Thus far only diagnostics that differentiate between "good" and "bad" have been considered. These can be thought of as two-level approaches. Diagnostic techniques can sometimes provide additional levels of diagnosis. Theoretically, a diagnostic may provide as many levels of diagnosis as physical states that exist for a component type. In other words, a four-level diagnostic for a component that has three states would not be useful. Such a diagnostic would be, by definition, highly inaccurate. Suppose, a three-level diagnostic is defined such that the diagnostic test classifies the components into classes C_1, C_2, or C_3, and that these classes correspond to component states α, β, and δ. The resulting probability of correctly classifying a component in condition α into the correct class C_1 is given by:

$$P(\alpha \mid C_1) = \frac{P(C_1 \mid \alpha) P(\alpha)}{P(C_1 \mid \alpha) P(\alpha) + P(C_1 \mid \beta) P(\beta) + P(C_1 \mid \delta) P(\delta)}. \quad \textbf{(6.10)}$$

As before, the denominator represents all possible ways that the diagnostic test can produce a classification of C_1. This can be rewritten into a more general form as,

$$P(\alpha \mid C_1) = \frac{P(C_1 \mid \alpha) P(\alpha)}{\sum_{Z \in Conditions} P(C_1 \mid Z) P(Z)}. \quad \textbf{(6.11)}$$

Clearly, as the number of diagnoses increases so do the number of ways the diagnostic can incorrectly classify a component.

6.5 MULTIPLE INDEPENDENT DIAGNOSTIC TESTS

In order to increase diagnostic accuracy, it is possible to employ multiple diagnostic tests. These tests may be either independent or dependent. This discussion focuses on the case of multiple independent diagnostic tests.

Suppose two diagnostic tests, T_1 and T_2, will be employed on a population of components containing a portion that are "bad" (B) and the remaining portion that are "good" (G). Probabilities can be defined that represent the probabilities of the two diagnostic tests correctly identifying the "good" and "bad" components as shown in Table 26.

Table 26: Definitions of probabilities for diagnostic tests T_1 and T_2.

Probability	Definition
$P(T_1 \mid G)$	The probability of correctly diagnosing a "good" component as good using T1.
$P(T_1 \mid B)$	The probability of incorrectly diagnosing a "bad" component as good using T1.
$P(\overline{T}_1 \mid G)$	The probability of incorrectly diagnosing a "good" component as bad using T1.
$P(\overline{T}_1 \mid B)$	The probability of correctly diagnosing a "bad" component as bad using T1.
$P(T_2 \mid G)$	The probability of correctly diagnosing a "good" component as good using T2.
$P(T_2 \mid B)$	The probability of incorrectly diagnosing a "bad" component as good using T2.
$P(\overline{T}_2 \mid G)$	The probability of incorrectly diagnosing a "good" component as bad using T2.
$P(\overline{T}_2 \mid B)$	The probability of correctly diagnosing a "bad" component as bad using T2.

Using the definitions in Table 26, it is then possible to compute the probability of a component actually being "good" or "bad" based on the diagnostic test results as follows:

$$P(G \mid T_1T_2) = \frac{P(T_1T_2 \mid G)P(G)}{P(T_1T_2)}$$

$$= \frac{P(T_1T_2 \mid G)P(G)}{P(T_1T_2 \mid G)P(G) + P(T_1T_2 \mid B)P(B)}, \tag{6.12}$$

where,

$P(G|T_1T_2)$ = Probability that a component is "good" given that T_1 and T_2 both diagnose it as good.

If the diagnostic tests are independent then (6.12) can be reduced to:

$$P(G \mid T_1T_2) = \frac{P(T_1 \mid G)P(T_2 \mid G)P(G)}{P(T_1 \mid G)P(T_2 \mid G)P(G) + P(T_1 \mid B)P(T_2 \mid B)P(B)}. \tag{6.13}$$

A similar equation may be written for the case when both diagnostics diagnose the component as bad when it is indeed "bad" as follows:

$$P(B \mid \overline{T}_1\overline{T}_2) = \frac{P(\overline{T}_1 \mid B)P(\overline{T}_2 \mid B)P(B)}{P(\overline{T}_1 \mid G)P(\overline{T}_2 \mid G)P(G) + P(\overline{T}_1 \mid B)P(\overline{T}_2 \mid B)P(B)}. \tag{6.14}$$

Similarly, the probabilities for the remaining combinations may be computed as:

$$P\left(G \mid \overline{T}_1 T_2\right) = \frac{P\left(\overline{T}_1 \mid G\right) P\left(T_2 \mid G\right) P(G)}{P\left(\overline{T}_1 \mid G\right) P\left(T_2 \mid G\right) P(G) + P\left(\overline{T}_1 \mid B\right) P\left(T_2 \mid B\right) P(B)},$$

$$P\left(G \mid T_1 \overline{T}_2\right) = \frac{P\left(T_1 \mid G\right) P\left(\overline{T}_2 \mid G\right) P(G)}{P\left(T_1 \mid G\right) P\left(\overline{T}_2 \mid G\right) P(G) + P\left(T_1 \mid B\right) P\left(\overline{T}_2 \mid B\right) P(B)},$$

$$P\left(G \mid \overline{T}_1 \overline{T}_2\right) = \frac{P\left(\overline{T}_1 \mid G\right) P\left(\overline{T}_2 \mid G\right) P(G)}{P\left(\overline{T}_1 \mid G\right) P\left(\overline{T}_2 \mid G\right) P(G) + P\left(\overline{T}_1 \mid B\right) P\left(\overline{T}_2 \mid B\right) P(B)},$$

$$P\left(B \mid \overline{T}_1 T_2\right) = \frac{P\left(\overline{T}_1 \mid B\right) P\left(T_2 \mid B\right) P(B)}{P\left(\overline{T}_1 \mid G\right) P\left(T_2 \mid G\right) P(G) + P\left(\overline{T}_1 \mid B\right) P\left(T_2 \mid B\right) P(B)},$$

$$P\left(B \mid T_1 \overline{T}_2\right) = \frac{P\left(T_1 \mid B\right) P\left(\overline{T}_2 \mid B\right) P(B)}{P\left(T_1 \mid G\right) P\left(\overline{T}_2 \mid G\right) P(G) + P\left(T_1 \mid B\right) P\left(\overline{T}_2 \mid B\right) P(B)},$$

$$P\left(B \mid T_1 T_2\right) = \frac{P\left(T_1 \mid B\right) P\left(T_2 \mid B\right) P(B)}{P\left(T_1 \mid G\right) P\left(T_2 \mid G\right) P(G) + P\left(T_1 \mid B\right) P\left(T_2 \mid B\right) P(B)},$$

(6.15)

where,

$P\left(G \mid \overline{T}_1 T_2\right)$ = Probability that a component is "good" given that T_1 diagnoses it as bad and T_2 diagnoses it as good,

$P\left(G \mid T_1 \overline{T}_2\right)$ = Probability that a component is "good" given that T_1 diagnoses it as good and T_2 diagnoses it as bad,

$P\left(G \mid \overline{T}_1 \overline{T}_2\right)$ = Probability that a component is "good" given that T_1 diagnoses it as bad and T_2 diagnoses it as bad,

$P\left(B \mid \overline{T}_1 T_2\right)$ = Probability that a component is "bad" given that T_1 diagnoses it as bad and T_2 diagnoses it as good,

$P\left(B \mid T_1\overline{T}_2\right)$ = Probability that a component is "bad" given that T_1 diagnoses it as good and T_2 diagnoses it as bad.

$P\left(B \mid T_1T_2\right)$ = Probability that a component is "bad" given that T_1 diagnoses it as good and T_2 diagnoses it as good,

Suppose the data shown in Table 27 are used to illustrate the use (6.13) thru (6.15).

Table 27: Definition of each variable for the two diagnostic example.

Variable	Value
$P(G)$	0.85
$P(B)$	0.15
$P(T_1 \mid G)$	0.75
$P(T_1 \mid B)$	0.25
$P(\overline{T}_1 \mid G)$	0.25
$P(\overline{T}_1 \mid B)$	0.75
$P(T_2 \mid G)$	0.90
$P(T_2 \mid B)$	0.10
$P(\overline{T}_2 \mid G)$	0.10
$P(\overline{T}_2 \mid B)$	0.90

Using the values in Table 27, the corresponding probabilities are shown in Table 28.

Table 28: Resulting probabilities for values presented in Table 27.

Variable	Value
$P(G \mid T_1T_2)$	0.993
$P(G \mid \overline{T}_1T_2)$	0.944
$P(G \mid T_1\overline{T}_2)$	0.944
$P(G \mid \overline{T}_1\overline{T}_2)$	0.173
$P(B \mid T_1T_2)$	0.006
$P(B \mid \overline{T}_1T_2)$	0.056
$P(B \mid T_1\overline{T}_2)$	0.056
$P(B \mid \overline{T}_1\overline{T}_2)$	0.827

Note that the data in Table 28 represent four complimentary pairs. This process may be extended to include more diagnostic tests. The general case of *d* independent diagnostic tests is:

$$P(G \mid T_1T_2...T_d) = \frac{P(G)\prod_{i=1}^{d} P(T_i \mid G)}{P(G)\prod_{i=1}^{d} P(T_i \mid G) + P(B)\prod_{i=1}^{d} P(T_i \mid B)}. \quad \textbf{(6.16)}$$

Using (6.16), any number of perfectly independent diagnostic tests may be combined together to increase the overall accuracy. As Table 28 shows, the accuracies increase by performing two diagnostic tests instead of just one. In cases where the diagnostic tests are not independent, then the gains in diagnostic accuracy will be less and could be as little as zero if the two diagnostic tests are completely dependent. In other words, nothing new is learned from the second diagnostic tests are completely dependent.

6.6 SUMMARY

This chapter has described the meaning of diagnostic accuracy and shown that there are two types of accuracies that are needed to develop a diagnostic program model: (1) overall accuracy and (2) condition-specific accuracy. In addition, two methods are available for computing these accuracies depending on the amount of information supplied by the diagnostic provider, namely: (1) "bad means fail" approach and (2) probabilistic approach. Of the two methods, the probabilistic approach is more representative of the way components fail in service. According to the example presented in Section 6.3.3, the probabilistic approach to diagnostic accuracy calculations clearly leads to higher accuracies of the "bad" groups.

CHAPTER 7: TECHNIQUES FOR ASSESSING DIAGNOSTIC TESTS

The primary objective of analyzing diagnostic data is to obtain the diagnostic accuracies as described in Chapter 6. Unfortunately, the interpretation of diagnostic testing results and the corresponding field data is quite challenging. This chapter describes several techniques that have been developed or adapted expressly for this purpose. All of these techniques have been used to assess diagnostic tests conducted in the field on underground cable systems.

7.1 PERFORMANCE RANKING

Performance ranking was developed as a means of evaluating the effectiveness of diagnostic testing by comparing the diagnostic data with service performance. This comparison provides a measure of the accuracy of the diagnostic. Ranking itself is a known procedure in statistics. In fact, a specialized version of the correlation coefficient exists for ranked data [29]. The key aspect of the development of this method is the process for generating the ranks (interpretation of diagnostic and service performance data) and the calculation of diagnostic accuracy from the ranks themselves.

Performance ranking is the only technique that looks at the entire spectrum of data from the best to the worst. In addition, it may be used with any diagnostic test was well as with data provided in any form.

The performance ranking technique is based on the generation of two distinct ranks, the performance rank and diagnostic rank, for each tested component. Each of these ranks is a number that gives the relative performance of each circuit compared to all other circuits in the group. There cannot be duplicate ranks within either rank type. Furthermore, all components must be assigned both a performance and diagnostic rank to be included in the analysis. In other words, if a test group consists of 10 components,

then there will be at most a single first rank, second rank, etc., for the performance rank as well as for the diagnostic rank.

The basic procedure can be summarized as follows:

1. Determine the performance rank using the available failure and component information.
2. Determine the diagnostic rank using the available diagnostic data and the component information.
3. Plot diagnostic rank versus performance rank.
4. Analyze the ranks using statistical techniques.

The concept of ranking the components is quite simple. However, with test groups containing more than a few components, there will likely be cases where the ranking criteria produce ties. As one of the requirements of this technique is to assign a single rank to each component, breaking these ties becomes critical. A hierarchy for the case of underground cables has been developed to address this issue for both ranks, each of which is discussed in detail in the following sections in conjunction with the steps outlined above.

7.1.1 Performance Rank

The performance rank is based on the failure data from either before or after testing. It is determined by comparing the failure rates (annual or cumulative) for all tested components with one another and ranking from worst (highest failure rate or shortest time to failure) to best (lowest failure rate or longest time to failure). The task can be complicated by the availability (or lack thereof) of failure information. For example, in the case of underground cables, failures are typically recorded for a complete "feeder," circuit that includes multiple cable segments. On the other hand, the diagnostic testing is generally performed on each cable segment. In these cases, one must ensure that the diagnostic data and failure data are at the same level of detail for the analysis to be valid.

Fortunately, this is not an issue for other components such as transformers, breakers, or poles. However, it is important to note that the ranking approach is able to cope with whatever detail is provided in the available data.

It must be noted that the performance rank is highly dependent on the amount of time that has elapsed since the diagnostic tests were carried out. Depending on the type of equipment and the current failure rate, it may take several years for enough data to be accumulated.

7.1.2 Diagnostic Rank

The diagnostic rank is far more complicated to determine than the performance rank because different diagnostic techniques provide their assessments in different ways. The data may be quantitative measures of the degradation that has occurred in the device or may simply be qualitative such as "good," "bad," or "okay." Furthermore, as with the performance rank this data may be as specific as by component or may include several components at once. Whatever the level of detail may be, it is necessary to evaluate the diagnostic data in the same groupings as the performance data.

Listed below are some examples of available underground cable diagnostic data that has been successfully analyzed using performance ranking. Note that this list is in no way exhaustive:

- Recommended sections of circuit for replacement.
- Partial discharge magnitude and count.
- Dielectric loss.
- Severity.

It must be emphasized that the only requirement for diagnostic data is that it be capable of providing some level of distinction between different circuits.

7.1.3 Ranking Tie Breaks

As mentioned above, it is common to see situations where ties can arise, especially in the case of the performance rank. These cases can be solved in a number of ways depending on the type of component. For example, in the case of cables, most ties may be dealt with by normalizing by the length of the segment. Other components such as breakers and transformers may be handled by considering age or even the number of exposures to fault currents. The key is to choose a characteristic that exhibits enough variability within the component population. It is also possible that multiple characteristics will be needed to break all the ties. For cables, the following hierarchy was developed based on the circuit information that is typically available at utilities:

- Circuit length: Average per unit length. Also, longer circuits should be more prone to failure so give higher rank to longer circuits.
- Number of accessories: More accessories lead to more opportunities for failure thus give higher rank to circuits with more accessories.
- Age: Older circuits receive a higher rank, as these are logically more prone to failure.
- Construction: Primarily, insulation type; however, this should also include jacketing, whether the cable was direct-buried or installed in conduit, and type of neutral.

7.1.4 Analyzing the Ranks

Once the two ranks have been computed, they may be analyzed either graphically (qualitatively) or statistically (quantitatively). In the former case, a plot of diagnostic rank versus performance rank is generated. A sample of such a plot is shown in Figure 55.

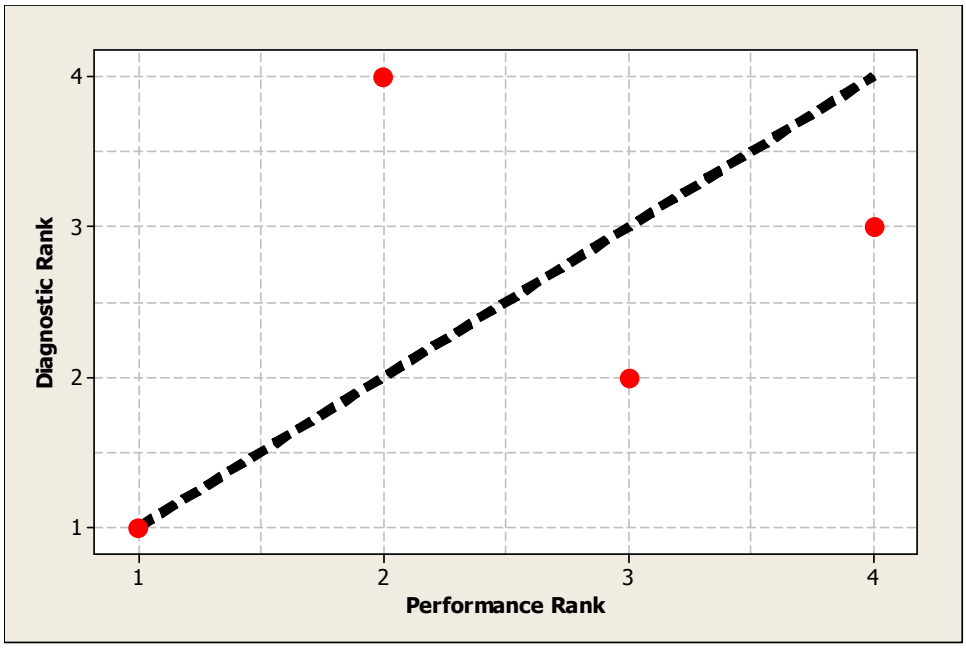


Figure 55: Sample performance ranking plot. Dots represent ranking data whereas the dashed line represents what would be a perfect correlation between performance and diagnostic ranks.

The interpretation of Figure 55 is as follows: Components in the lower left corner are the worst performers (highest failure rate and classified as "bad" by the diagnostic test) while the upper right corner contains the best performing components (low failure rate and classified as "good" by the diagnostic test). This is further illustrated in Figure 56. The closer the ranking points are to the hypothetical perfection line (dashed line in Figure 55 and Figure 56), the more accurate the diagnostic was at evaluating the particular group of cables.

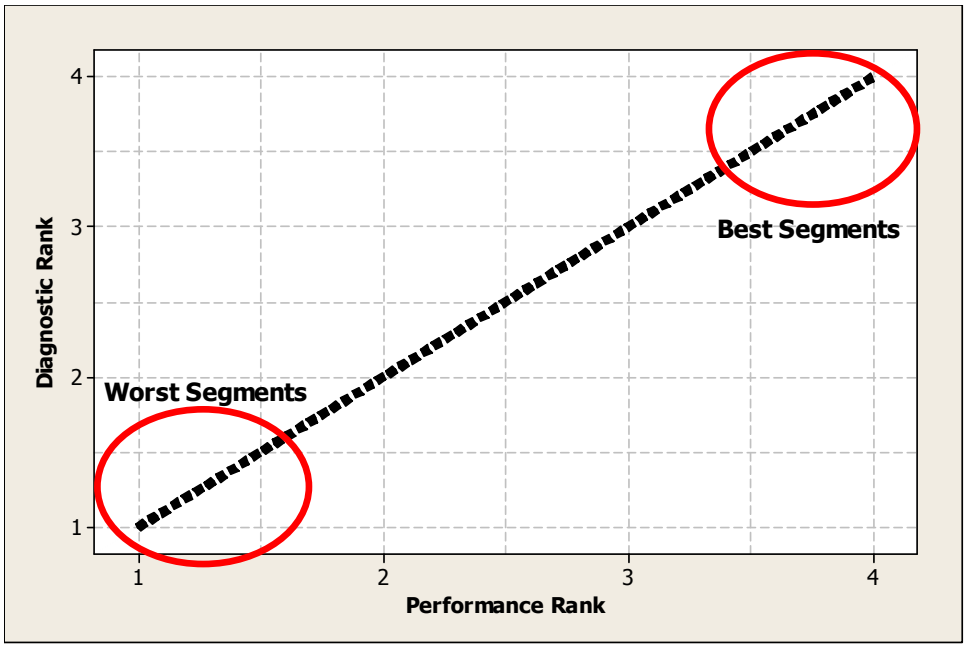


Figure 56: Annotated performance rank plot showing the meaning and interpretation of the "perfect assessment" line.

The dashed line can be thought of as perfect correlation between the performance and diagnostic ranks. Therefore, the obvious statistical approach is to examine the Pearson correlation coefficient between the two ranks as,

$$r_{DP} = \frac{n\sum D_i P_i - \sum D_i \sum P_i}{\sqrt{n\sum D_i^2 - \left(\sum D_i\right)^2}\sqrt{n\sum P_i^2 - \left(\sum P_i\right)^2}}, \qquad \textbf{(7.1)}$$

where,

r_{DP} = Pearson correlation coefficient,

n = number of samples,

$D_i = i^{th}$ Diagnostic Rank,

$P_i = i^{th}$ Performance Rank.

For the example shown in Figure 55, the Pearson correlation coefficient is 0.40. In addition, the correlation coefficient carries with it a specified level of significance (p

value) based on the correlation value and number of samples. This *p*-value represents the probability of obtaining the observed correlation coefficient at random given the same number of samples. In other words,

$$p_n = P\left(r_{DP} \mid n \text{ samples}\right). \quad \textbf{(7.2)}$$

The resulting *p*-value for the example in Figure 55 is greater than 0.1 and indicating that the obtained correlation coefficient would occur randomly with probability greater than 0.1. Typically, *p*-values should be less than 0.05 for the correlation to be considered significant.

7.1.5 Diagnostic Accuracy based on Performance Ranking

The Performance Ranking technique can be used as a basis for establishing the overall diagnostic accuracy for the diagnostic technique employed on a particular utility system. An example will be used (shown in Figure 57) to demonstrate how this process may be performed.

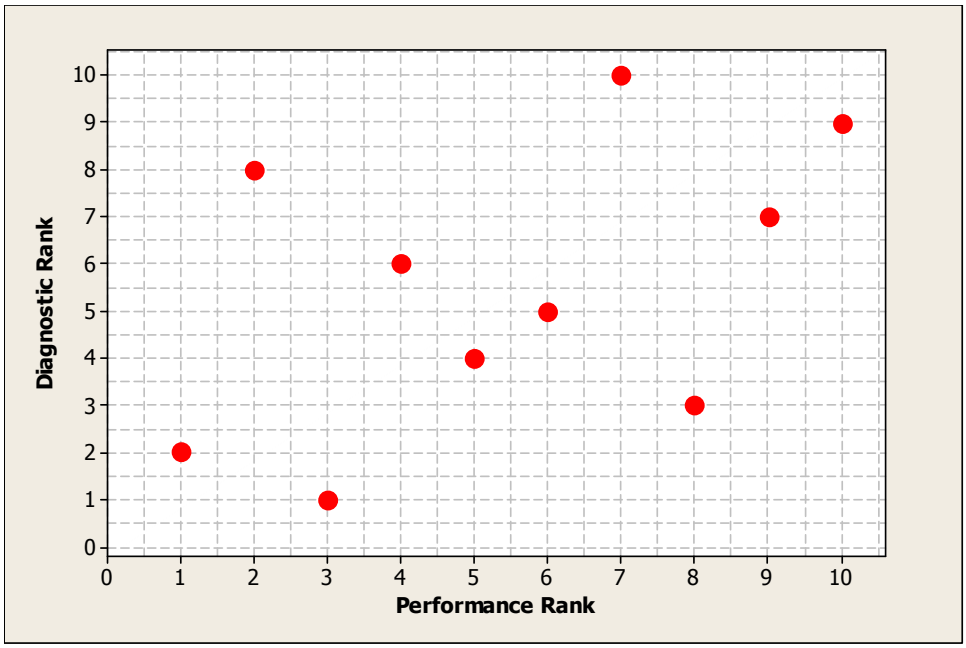


Figure 57: Example performance ranking plot for a diagnostic program performed on a population of 10 components.

The process for obtaining the overall diagnostic accuracy is described in the following sections.

7.1.5.1 Step 1: Separating Diagnostic Ranks

The first step in this procedure is to separate the diagnostic ranks into two classes: "good" and "bad" or "action required" and "no action required." This process must be done according to the type of diagnostic data that is available and the chosen method of determining the rank. As described above, different diagnostics provide their assessments in different ways but all of them classify components, either directly or indirectly, into "good" and "bad." It is entirely possible that all the ranks could be classified by the diagnostic as only "good" or only "bad." This is not an issue for computing the overall diagnostic accuracy. Although, in practice such a scenario would indicate a less than ideal choice of component population as discussed in Part III.

To make the classification, one must have an understanding what the diagnostic data mean. Once the two ranks are identified that correspond to the boundary between "good" and "bad" a line is added to the performance ranking plot as shown, for example, in Figure 58.

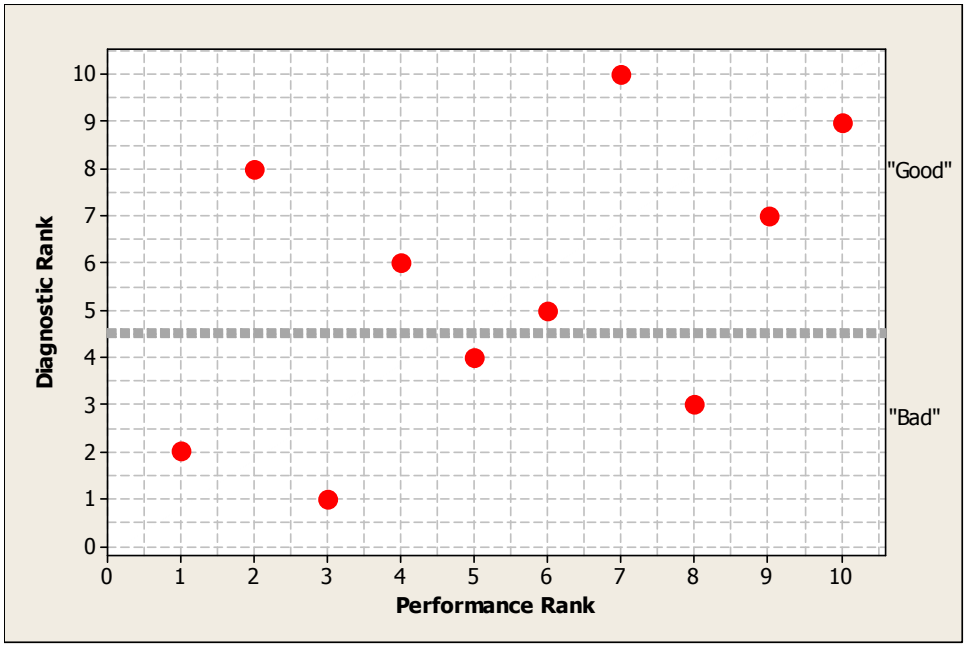


Figure 58: Performance ranking plot with boundary between diagnostic ranks corresponding to "good" (top) and "bad" (bottom).

7.1.5.2 Step 2: Separating Performance Ranks

The second step in the procedure requires that the boundary between the performance ranks that correspond to components that did not fail in service and those components that did fail in service. The data are far simpler to interpret since the only information needed is whether or not the component experienced a service failure. As in step 1, once the boundary ranks are determined a line is added to the performance ranking plot as shown in Figure 59.

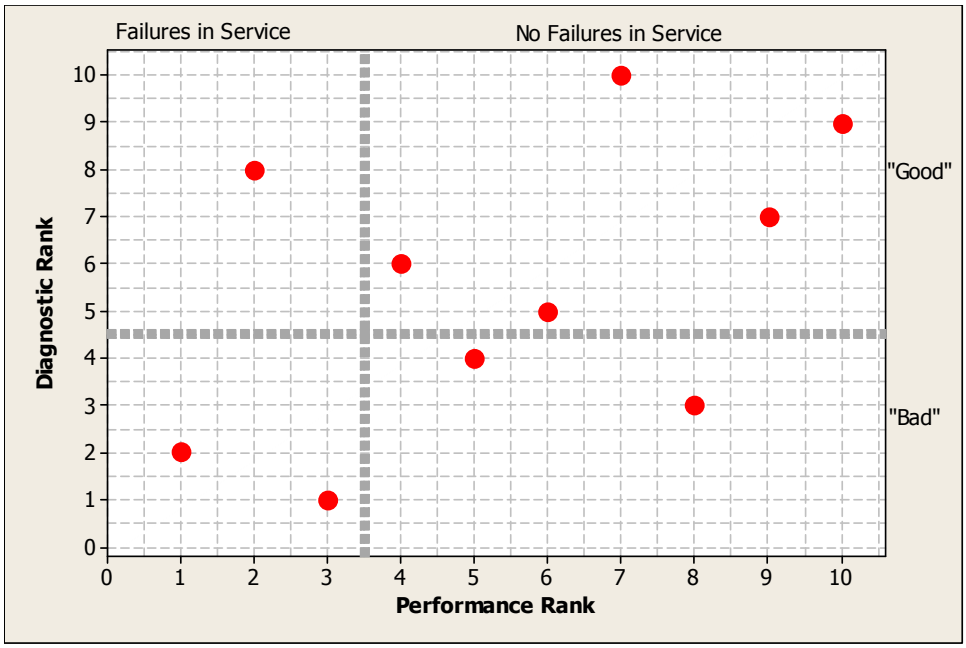


Figure 59: Performance ranking plot with boundary between ranks corresponding to "failures in service" (left) and "no failures in service" (right).

7.1.5.3 Step 3: Compute the Overall Diagnostic Accuracy

The third step in the procedure is to compute the overall diagnostic accuracy using the four unequal quadrants identified as A_1 thru A_4 in Figure 60. Using these variables to represent the number of data points contained in their respective quadrants produces the data shown in Table 29. Note that Table 29 also lists the specific definition for each quadrant. For example, quadrant A_1 contains the components that were diagnosed as "bad" and then subsequently failed in service.

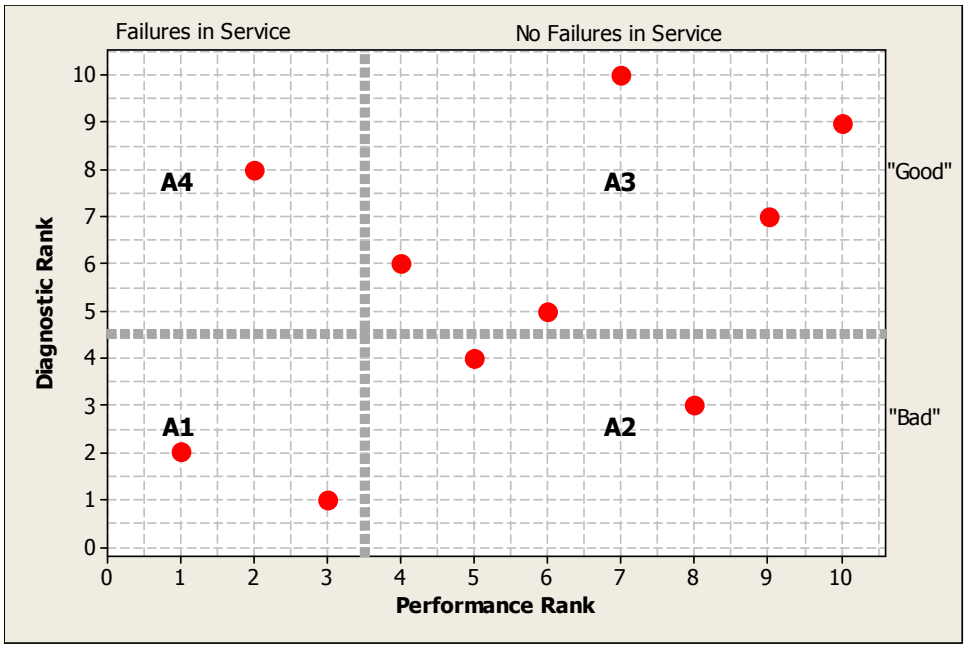


Figure 60: Sample performance ranking plot with regions defined for the calculation of diagnostic accuracies.

Table 29: Definitions of the quadrants shown in Figure 60.

Quadrant	Description	Components [#]
A_1	Components diagnosed as "bad" and failed in service	2
A_2	Components diagnosed as "bad" and did not fail in service	2
A_3	Components diagnosed as "good" and did not fail in service	5
A_4	Components diagnosed as "good" and failed in service	1

As described in Section 6.3, two methods are available for computing the overall diagnostic accuracy. The calculation for each method is shown below:

Method 1

Equation (7.3) shows the computation steps. Essentially, A_1 and A_3 represent the number of "bad" components and "good" components that were correctly assessed by the diagnostic, respectively. The assumption is that all "bad" components must fail by the

time of analysis while all "good" components must not fail. Using this assumption, the summation of these two quadrants provides the total number of correct assessments:

$$\begin{aligned}\text{Overall Diagnostic Accuracy} &= \frac{A_1 + A_3}{\sum_{i=1}^{4} A_i} \times 100\% \\ &= \frac{2+5}{2+2+5+1} \times 100\% \\ &= 70\%.\end{aligned} \tag{7.3}$$

The resulting overall diagnostic accuracy is 70% for this example.

Method 2

This second method is somewhat more complicated as it requires knowledge of the probability of failure for the components diagnosed as "bad." For the example, suppose data are available that would allow one to determine the probability of failure for a component diagnosed as "bad" and one diagnosed as "good" at the present time, denote these probabilities as P_b and P_g, respectively. Given this probability, the resulting overall diagnostic accuracy calculation is:

$$\text{Correct "Good" Diagnoses} = N_g,$$

$$\text{Correct "Bad" Diagnoses} = N_b,$$

$$N_g = \begin{Bmatrix} A_3 + 2A_4 - P_g\left(A_3 + A_4\right) & A_4 < P_g\left(A_3 + A_4\right) \\ A_3 + P_g\left(A_3 + A_4\right) & A_4 > P_g\left(A_3 + A_4\right) \\ A_3 + A_4 & A_4 = P_g\left(A_3 + A_4\right) \end{Bmatrix},$$

$$N_b = \begin{Bmatrix} 2A_1 + A_2 - P_b\left(A_1 + A_2\right) & A_1 < P_b\left(A_1 + A_2\right) \\ A_2 + P_b\left(A_1 + A_2\right) & A_1 > P_b\left(A_1 + A_2\right) \\ A_1 + A_2 & A_1 = P_b\left(A_1 + A_2\right) \end{Bmatrix}, \quad \textbf{(7.4)}$$

$$\text{Overall Accuracy} = \frac{N_g + N_b}{\sum_{i=1}^{4} A_i} \times 100\%.$$

Note that quadrants A_2 and A_4 are now included in the calculation since P_b and P_g indicate that there should be data points in these two quadrants. Therefore, the calculation must consider the correct distribution between "bad" components that did fail and those that did not. The same applies for the components diagnosed as "good." If, as an example, the assumption is made that $P_b = 0.75$ and $P_g = 0$. Then the calculation of the overall diagnostic accuracy using (7.4) produces an accuracy of 80%. This is slightly higher than the overall accuracy computed using Method 1.

Of the two available methods for computing diagnostic accuracy Method 2 is preferred since it is more representative of the way components degrade and fail. Unfortunately, this method requires additional data that may or may not be available. In such cases, Method 1 is applicable and does provide useful information on the overall diagnostic accuracy from performance ranking data.

7.2 DIAGNOSTIC OUTCOME MAPPING

Diagnostic outcome mapping (DOM) uses failure and testing data to evaluate whether or not changes in reliability are coincident with diagnostic testing applications (as well as the activity called for by the test results). It is also a means of evaluating the quantities of testing and the current maintenance practices.

7.2.1 DOM Basics

DOM relies on the Crow-AMSAA technique [14]. Crow-AMSAA is a plotting technique that plots cumulative failures versus time on log-log scales. An example of such a plot is shown in Figure 61.

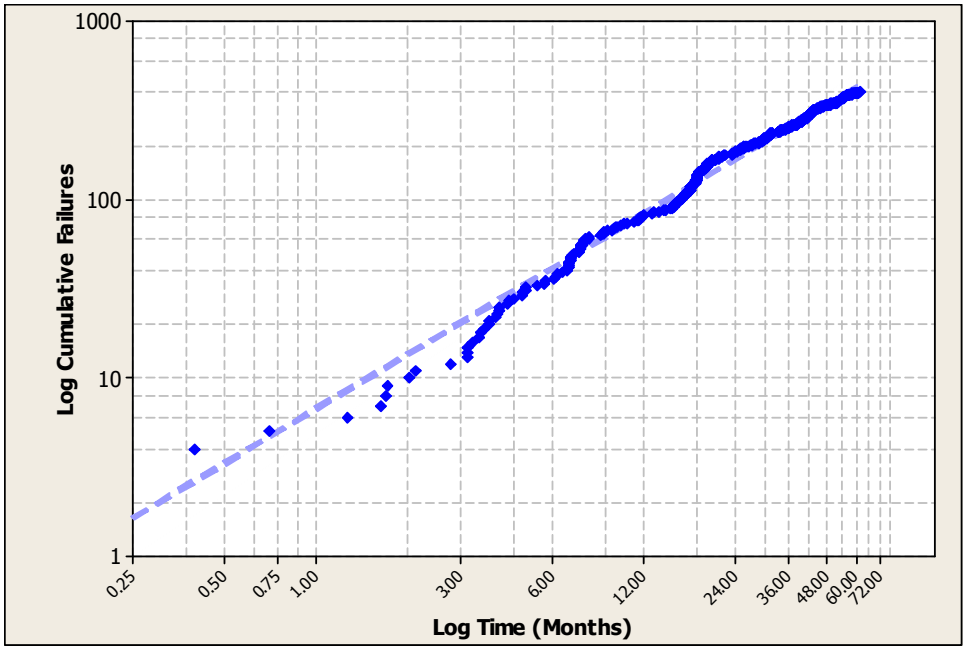


Figure 61: Sample Crow-AMSAA (Cumulative Failures versus Time).

The instantaneous failure rate is found by computing the slope, or gradient, of the curve. A decreasing gradient indicates the failure rate is decreasing, while an increasing gradient corresponds to an increasing failure rate. By adding the testing events to the

same plot, it is possible to see the effect a test program (diagnostic testing plus required action) has on the reliability.

7.2.2 DOM Examples

Figure 62 shows how a reduction in failure rate would appear following testing and action in an underground cable system. Note that multiple testing and action events are recorded as this circuit was tested in multiple segments while the failure records are only specified for the group of segments. Figure 62 shows that the failure following the testing and action events occurs later that would have been predicted by the line fitting the previous three failures.

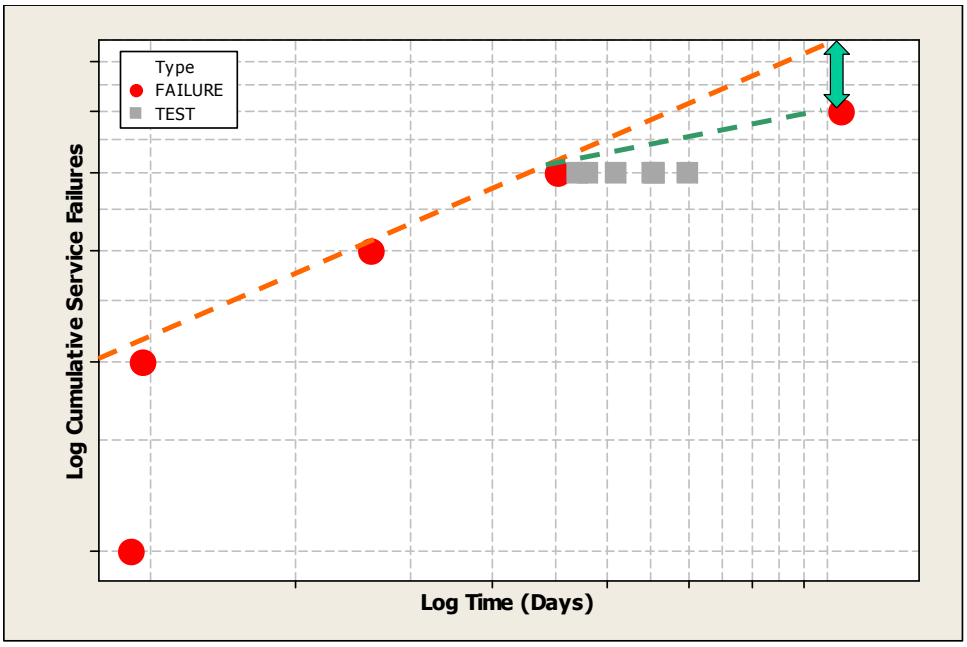


Figure 62: Sample DOM plot for decreasing failure rate scenario. Failures are shown as dots while testing/action events are shown as squares. Note that this figure corresponds to one entire feeder so multiple testing events were done on different segments of this feeder.

Figure 63 shows the same concept as Figure 62 except this example shows the testing program has not yet made an impact on the failure rate.

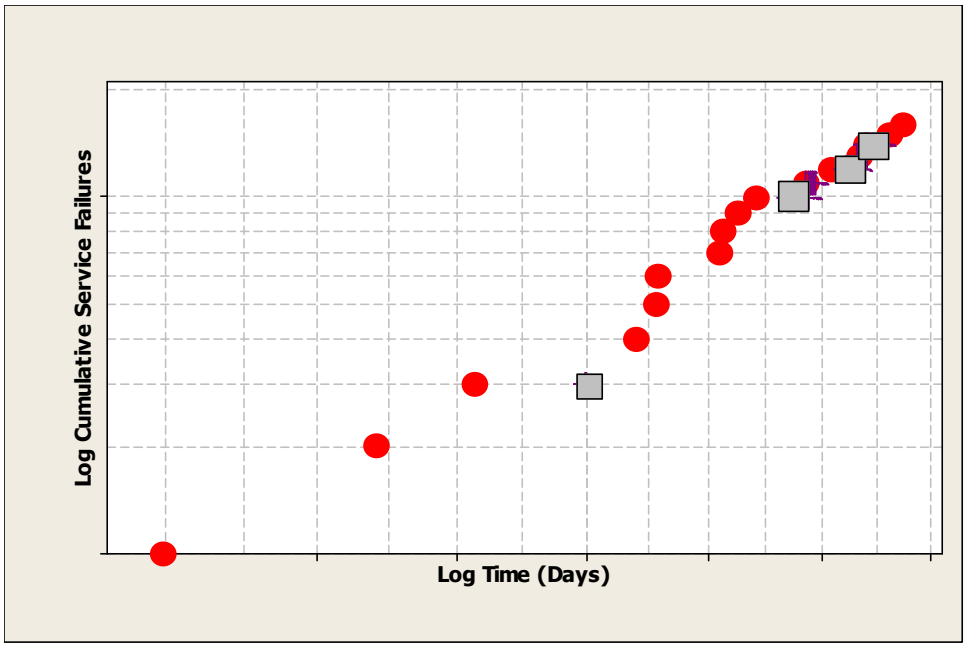


Figure 63: Sample DOM plot for "no change" scenario. Failures are shown as dots while testing/action events are shown as squares.

Using an outcome map, one can easily show if improvements in reliability are, indeed, the result of a diagnostic testing and action program. Over a long enough time period, the annual gradients can be examined to see whether or not reliability has improved. Figure 64 shows an example of a multi-year program. Note that after the third year, the gradients decrease until at year six it is 40% lower than at the start. This indicates a reduction in failure rate of 40% as compared to Year 0.

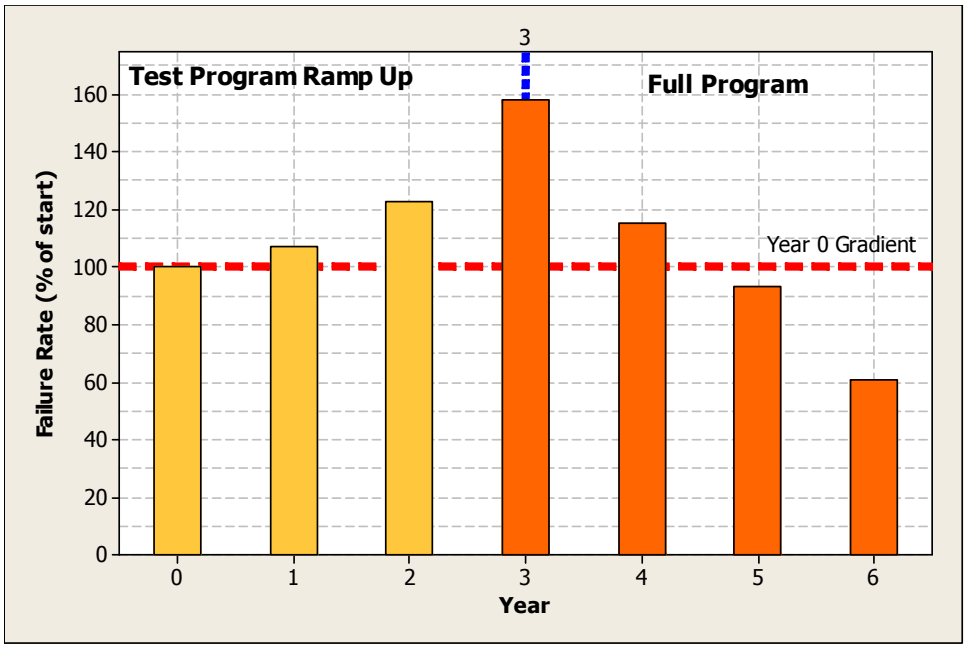


Figure 64: Sample gradients from am multi-year diagnostic testing and action program.

7.3 WEIBULL ANALYSIS

Weibull analysis is a technique that looks primarily at the time to failure (TTF) for different component groups. Using this TTF data, one can fit either a two or three-parameter Weibull distribution. Any distribution may be used but the Weibull distribution is particularly useful because of its correspondence to the well-known (and intuitive) bathtub curve (shown in Figure 65) [30]-[31].

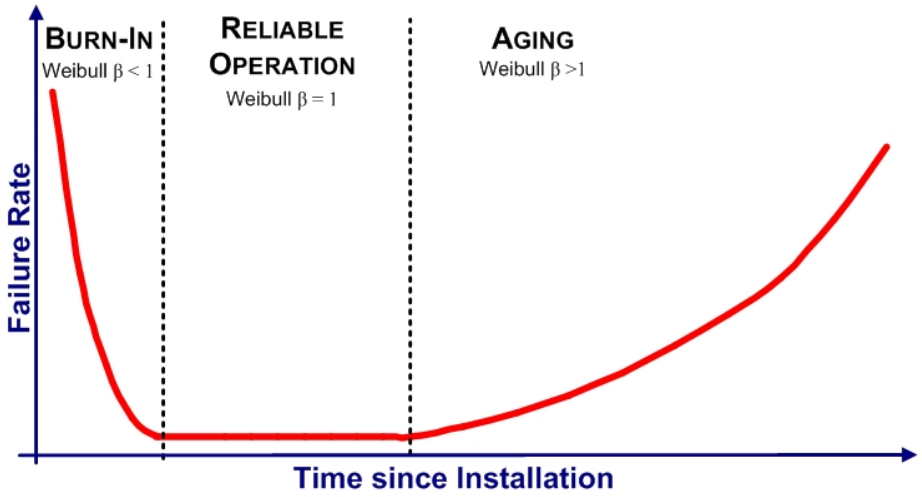


Figure 65: Sample Weibull bathtub curve with regions of operation shown and their corresponding Weibull gradients.

Figure 65 shows three regions, each of which corresponds to different failure modes. The first is the "early failures" region otherwise known as *infant mortality* or *burn-in*. This region corresponds to a Weibull gradient (β) that is less than one and would include failures that are caused by workmanship or installation errors. The *infant mortality* stage represents only a small percentage of the total operational life of the component, after which, the component will operate normally with a low failure rate (purely random failures) as represented by the *reliable operation* region. This region corresponds to a Weibull gradient of one. Failures that occur during this period are not related to aging, in fact, this is the most reliable period in a component's lifetime. In practice, components should spend the majority of their lifetimes in the *reliable operation* region. On the other hand, as time passes the components move from *reliable operation* into the *aging* region. This last region is characterized by a Weibull gradient that is greater than one and is the period during which most components will fail.

7.3.1 Input Data

In addition to the intuitive understanding that comes from the "bathtub" curve, Weibull analysis can be used with all the available data. As mentioned above, Weibull analysis focuses on the times to failure for the different components.

7.3.2 Uses for Weibull Analysis

There are two main uses for Weibull Analysis in the context of diagnostic testing: (1) validation of the recommendations made based on diagnostic measurements and (2) evaluation of diagnostic procedures themselves.

7.3.2.1 Evaluating Diagnostic Testing Recommendations

Weibull analysis is a useful diagnostic validation tool since, like performance ranking, it relates diagnostic assessments to service performance. The procedure focuses on the times to failure following testing for each assessment class (i.e. "good" and "bad") provided by the diagnostic test. By definition, the subpopulation of components designated as "bad" by the diagnostic test should have shorter times to failure than the subpopulation of "good" components. Consequently, the times to failure for the "bad" components will follow a Weibull distribution with a higher gradient (β) than the Weibull that corresponds to the "good" component times to failure. Ideally, this gradient will be greater than one, however, it need only be greater than the gradient from the "good" subpopulation to show the diagnostic test delivered useful information.

An example is shown in Figure 66 in which a diagnostic test has been performed on a population and recommendations made based on those measurements.

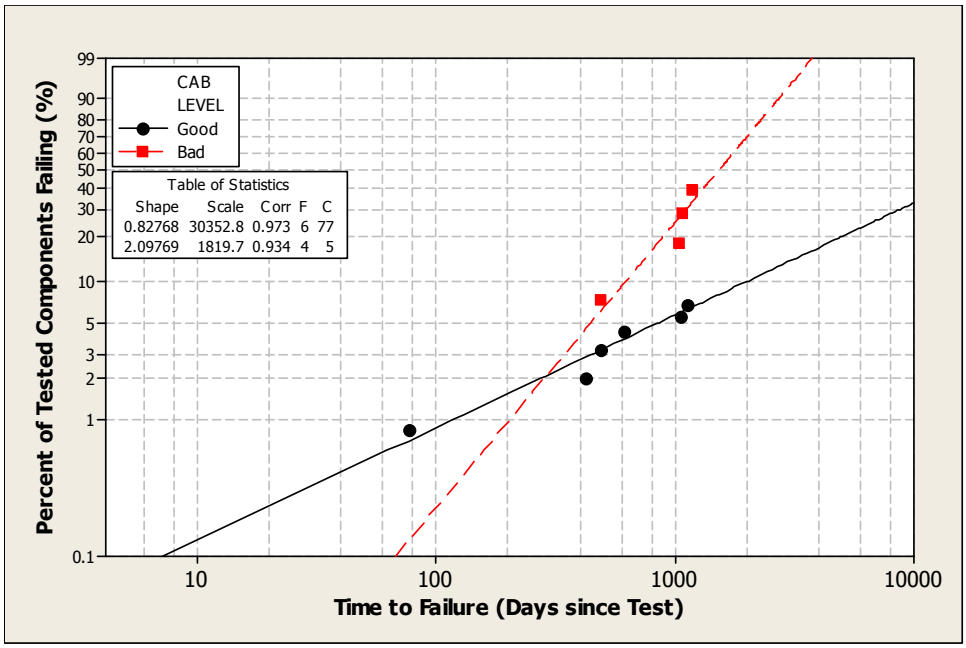


Figure 66: Weibull probability plot showing the times to failure for components diagnosed as "good" (●) and "bad" (■). The x-axis corresponds to the time to failure in days since test while the y-axis represents the percent of <u>tested</u> components failing.

The Weibull gradients are 0.83 and 2.10 for components diagnosed as "good" and "bad", respectively. In this case, the diagnostic provides sensible results as the "bad" components are clearly in the *aging* region of the bathtub and the "good" components are near the *reliable operation* region. Note that the "good" group does experience a failure before the "bad" group but this is because the "good" group is eight times larger. One should not make conclusions solely based on which group fails first but on the performance of the entire population.

7.3.2.2 Evaluating Diagnostic Procedures

Using the same concept as for the evaluation of diagnostic testing recommendations, Weibull Analysis can also be used to evaluate the performance of different diagnostic protocols.

In practice, the primary interest in terms of performance from a diagnostic program is its effect on the reliability of the tested population. Different diagnostic protocols generally lead to different service performance as well. Figure 67 shows the Weibull curves for two such protocols.

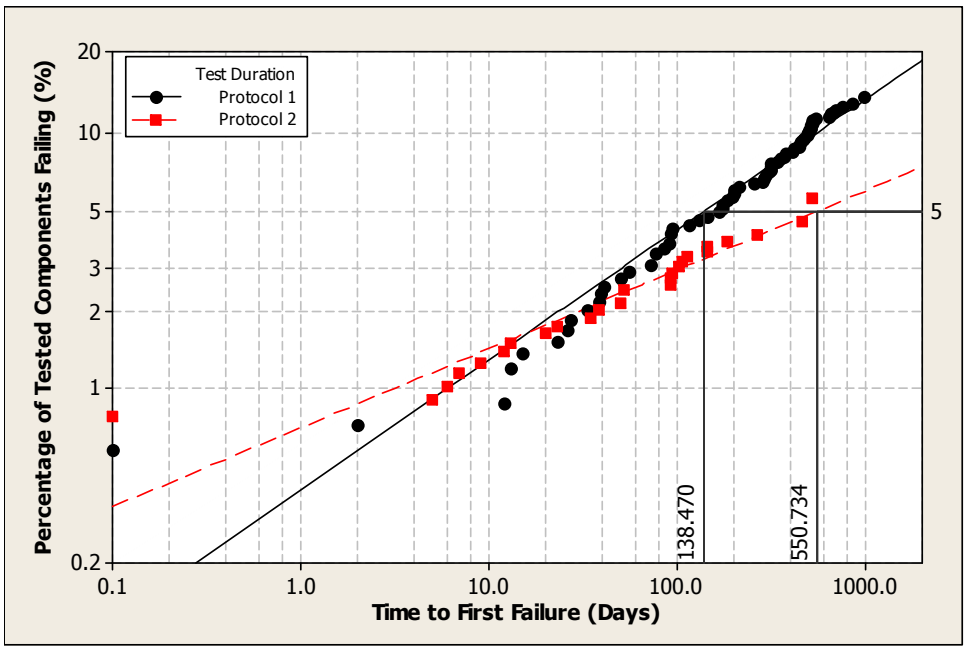


Figure 67: Weibull probability plot showing times to first failure for 15 (●) and 30 (■) minute tests. The x-axis corresponds to the time to failure in days since test while the y-axis represents the percent of <u>tested</u> components failing.

One way to compare the two test protocols is by examining the expected time for the same percentage of the population to fail. In the example shown in Figure 67, by choosing the time until 5% of the tested components fail, it is clear that the times until the utility should expect to see this percent of its population fail are 138 and 550 days for Protocols 1 and 2, respectively. Clearly, Protocol 2 produces more reliable service performance than Protocol 1.

7.4 SURVIVOR ANALYSIS

Survivor Analysis is another technique used to examine failure data. As part of this research, it has been widely used in the analysis of data from stress type tests otherwise known as withstand tests. This diagnostic is of particular interest to utilities as it is the only diagnostic test that does not require significant interpretation to be effectively understood.

Survivor analysis looks at the entire population of components under test and generates a function showing the failures during the test as a function of time. In mathematical terms, suppose the times on test for a group of components correspond to a random variable, x, and this random variable can be represented as a probability density function (PDF) called p_x. The PDF can then be integrated to give the cumulative distribution function (CDF), P_x. The survivor function is then given as [11],

$$S_x = 1 - P_x, \tag{7.5}$$

where,

S_x = Survivor function for random variable x.

The survivor function can be expressed either as a parametric distribution, such as Gaussian or Weibull, or may be computed using non-parametric techniques such as the Kaplan-Meier estimator [32].

7.4.1 Survivor Analysis Applications

The survivor technique has been used extensively to determine appropriate duration for "stress" type tests in which the components are stressed beyond their normal operating regions for a period of time. Such a test procedure is designed to drive weak ("bad") components to failure during the test so they can be replaced under controlled conditions. In the electric utility industry these tests are known as withstand tests where

the applied stress is in the form of elevated voltage. Among the main issues of these test programs are the voltage and time duration of the test. The duration of the test is very important as too short a time can cause the test to miss "bad" components while too long a time may cause unnecessary damage to components that were actually "good".

The following sections demonstrate the application of the survivor technique on high voltage withstand tests presently employed on underground cable systems. Both non-parametric and parametric approaches are demonstrated.

7.4.2 Non-Parametric Survivor Function

The construction of a non-parametric survivor curve is relatively straightforward once the data is suitably organized. Figure 68 shows the compilation of withstand test data for US utilities that use different combinations of test voltage and test duration.

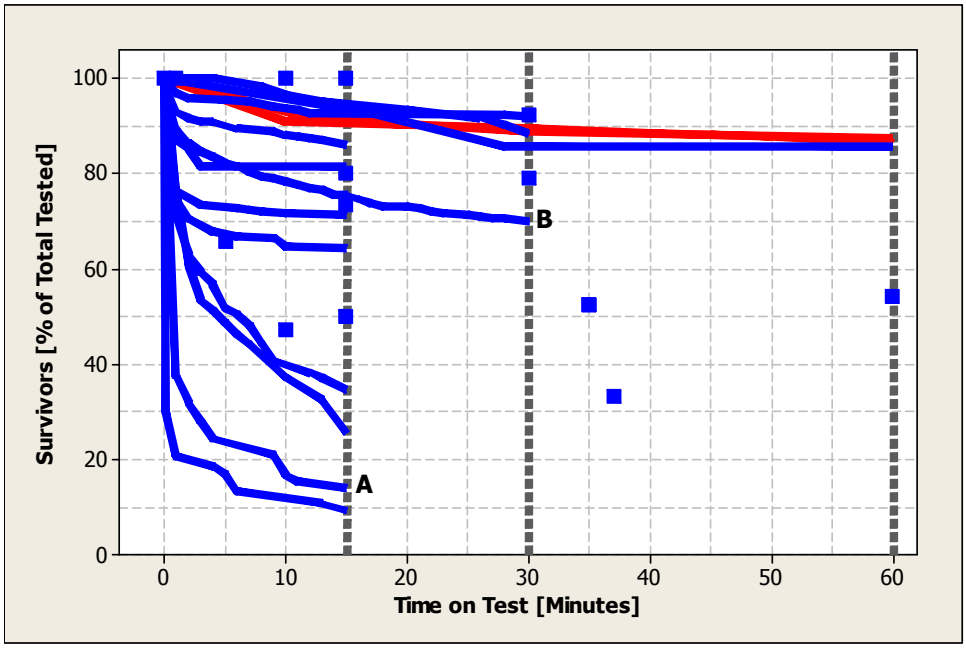


Figure 68: Example of Survivor Analysis for different utilities employing a high voltage "stress" test to their underground cable systems. The x-axis represents the time on test and the y-axis the corresponding percent of tested components that had not failed up until that time.

The shapes of these curves are related to the test protocol as well as the physical makeup of the systems under test. In all cases, though, the curves appear to flatten towards an asymptote. The existence of this asymptote implies that there is a point at which no further testing is needed since all failures that were going to occur have already done so. This is intuitive since as the curves flatten more and more there is a decrease in the number of failures that occur for each additional minute of testing. In practice, there is a point at which the value of finding the next failure is outweighed by the time investment to do so. Therefore, most utilities test for shorter times than would be needed to reach the asymptote. The survivor curves allow one to estimate the risk (the number of pending failures that do not occur during the test) of ending the test at any particular time.

7.4.2.1 Censoring

The greatest value of the survivor technique is that the method does not exclude data. Traditional analyses of withstand type data only examine the failures that occur during the test while completely ignoring the segments or components that did not fail. As Figure 68 shows, even the 60 minute long test did not cause all the components to fail during testing. In fact, only 15% of the tested components failed during this test program. This method has even been used by standards bodies in providing recommendations to the industry, such as in IEEE Std. 400.2 [38]. The survivor technique, on the other hand, is able to include these "un-failed" components by treating them as censors or suspensions.

Censoring is a technique that allows one to include data for which only boundaries on their values are known. For example, in the 60 minute test mentioned above, only 15% of the tested population failed during the test while 85% did not fail. It is not known exactly when the 85% of the population would have failed but it is sometime longer than 60 minutes. In this case, a lower bound on the times to failure for these components can be identified. This is often termed as "right censoring" since the unknown true times to

failure are greater than (to the right on a number line) of a known point [11]. Cases may also occur where it is possible to assign a maximum or upper bound on the failure time. This is termed as "left censoring" since the actual failure time is said to occur before or no later than the specified censored time.

The concept of censoring is vitally important to the analysis of any failure data and will be revisited on several occasions throughout the remainder of this document.

7.4.2.2 Asymptote

With the data shown in Figure 68, one can begin to analyze the different testing procedures. For example, curves A and B show different results for tests lasting 15 minutes and 30 minutes, respectively. The question one may ask is: Is 15 minutes enough time? To answer this question, it is possible to examine the gradient of curve A. If the test is long enough, then the gradient should be close to zero, in other words, completely flat. This would indicate that the test has found all "weak" locations and that no further testing should be necessary. As the gradient is not close to zero, it is clear that the 15 minute test is not long enough as the each additional minute of testing is able to locate another "weak" location. The question then becomes: How close to zero must the gradient be? Engineering judgment may be used to answer this question. However, one may also construct a model that would predict the expected time on test at which the curve would finally settle. A simple double exponential function of the form shown in (7.6) may be used.

$$y = a_1 e^{-\lambda_1 t} + a_2 e^{-\lambda_2 t}, \qquad \textbf{(7.6)}$$

where,

a_1, a_2, λ_1, λ_2 = Model parameters to be determined from the available data.

The model in (7.6) has been used to fit curve B in Figure 68. The results of this fit are shown in Figure 69. This figure shows that for this 30 minute duration test the settling

point for the resulting failures on test occurs between 80 and 100 minutes. Note that this model is consistent only with the exponential distribution, which is a special case of the Weibull distribution, so the ability of this approximation is limited at best.

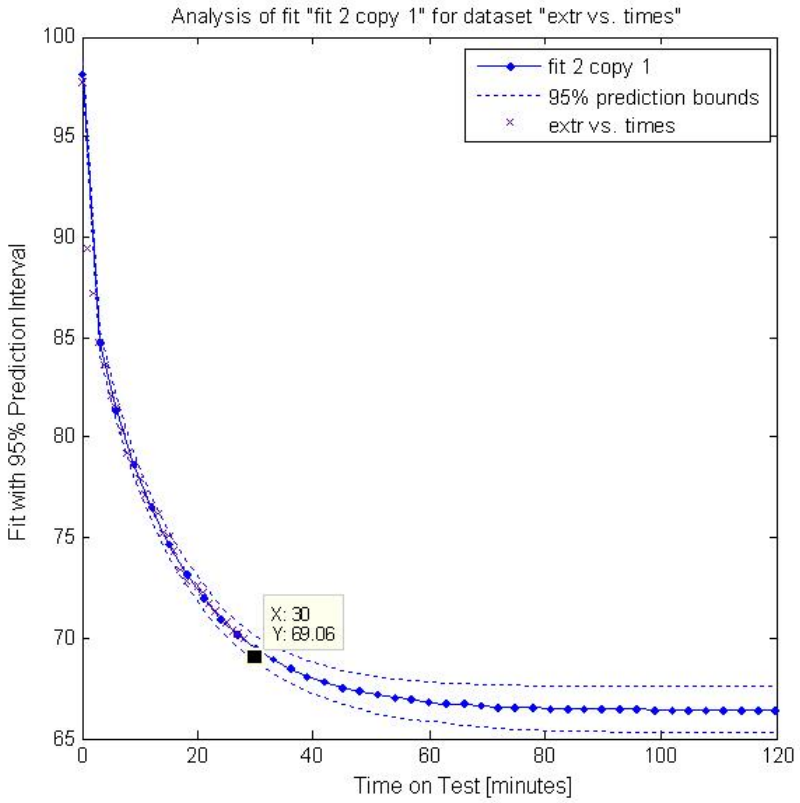


Figure 69: Sample double exponential fit for Curve B in Figure 68.

Using (7.6) as the model, the settling point can be calculated by taking the derivative and equating to zero as,

$$\frac{\partial y}{\partial t} = -a_1\lambda_1 e^{-\lambda_1 t} - a_2\lambda_2 e^{-\lambda_2 t} = 0$$
$$t = \frac{a_1\lambda_1}{a_2\lambda_2\left(\lambda_2 - \lambda_1\right)}. \qquad \textbf{(7.7)}$$

This settling point provides a measure of the number of weak locations that are being left in service by stopping the test at a shorter time as dictated by engineering judgment.

Based, on this information, a decision can be made as what is a tolerable number of missed weak locations.

7.4.3 Parametric Survivor Function

Weibull Analysis can also be used as the basis for a parametric approach to Survivor Analysis. As Section 7.4.2.2 shows, there is an interest in fitting a model to the survivor function so that predictions and further analysis made be performed. Using a parametric approach to generating the survivor curves fulfills both objectives. Using the data shown for curve B in Figure 68, Figure 70 shows the resulting Weibull probability curve. This curve is equivalent to P_x shown in (7.5) and, thus, represents the probability of failure at a given time.

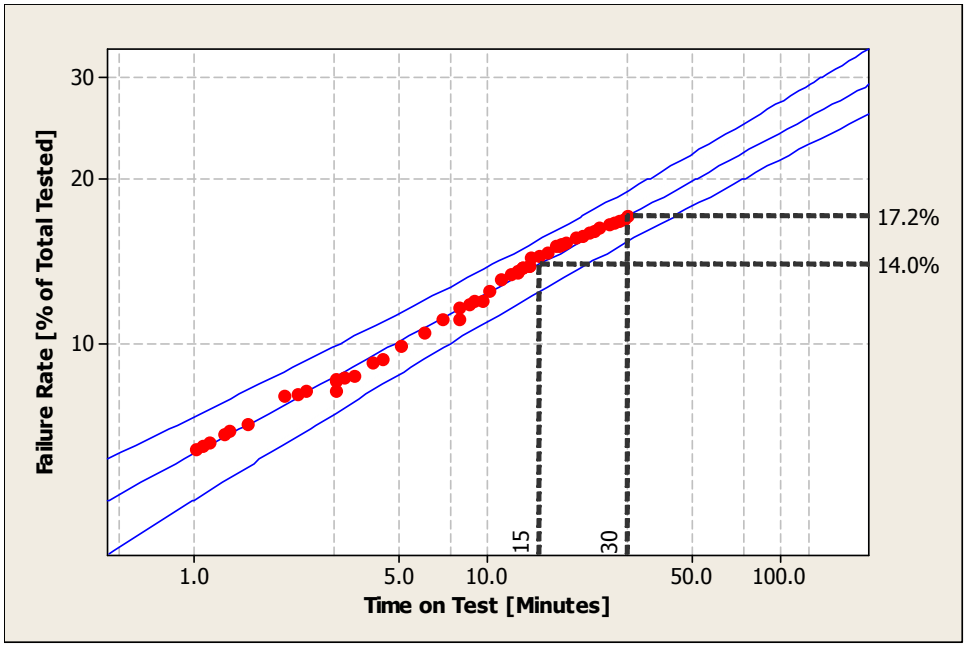


Figure 70: Weibull probability plot of times to failure for underground cable for a utility that employs a high voltage "stress" test. The x-axis represents the time on test and the y-axis the corresponding percent of tested components that failed by that time.

As Figure 70 shows, the 30 minute data fit very well to a single Weibull distribution since the data behave linearly and lie well inside of the 95% confidence bounds. Furthermore, since the data does not appear to contain any sharp knees (changes in slope) the choice of a single Weibull is likely to be appropriate. This data can be reformulated into the survivor form of Figure 68 and is shown in Figure 71.

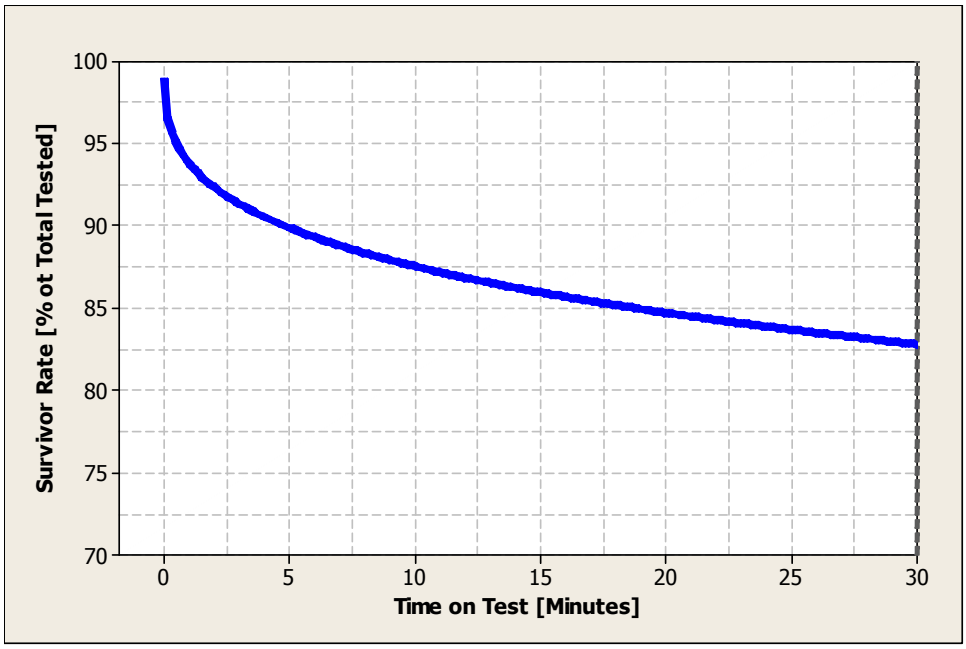


Figure 71: Parametric survivor curve that corresponds to data plotted in Figure 70.

Both Figure 70 and Figure 71 have the potential to be used to predict the survivor rates for longer testing times than 30 minutes (assuming all other test conditions remain the same). However, with any prediction there is a doubt as to its validity. As a test of this method, it is possible to reformulate the available 30 minute test data into a shorter test program of 10 minutes. Predictions for 15 and 30 minute survivor rates may then be made based on the 10 minute data. These predictions can be compared with the actual performance of the complete 30 minute dataset to verify the validity of this prediction

method. Figure 72 shows the failure rates for the 10 minute dataset and the corresponding predictions for 15 and 30 minutes.

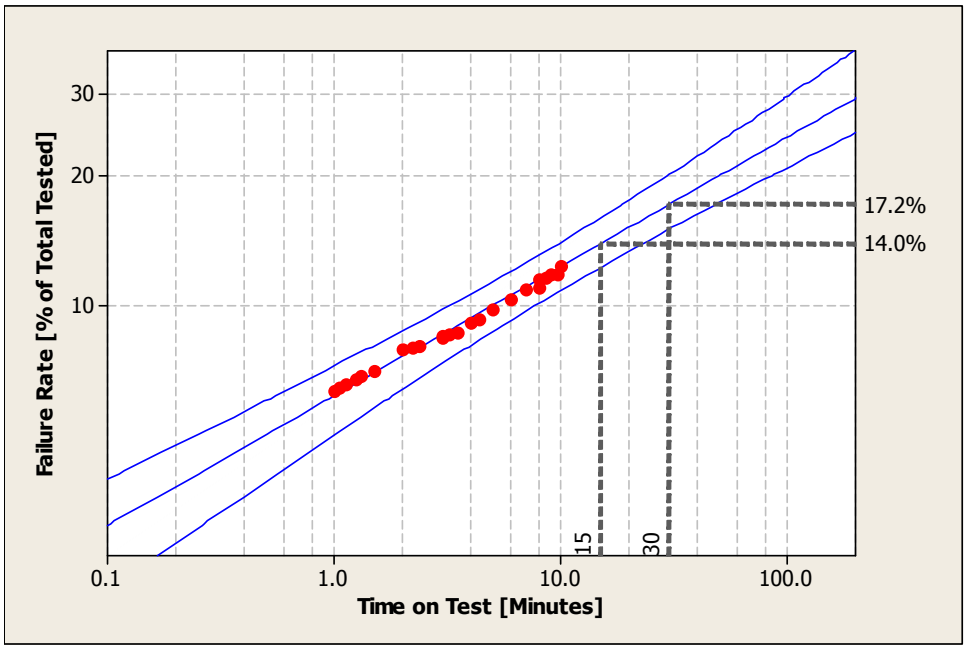


Figure 72: Example of subset of 30 minute dataset (10 minutes worth) being used to predict failure percentages for 15 and 30 minute tests.

Table 30 compares the predictions based on the 10 minute test dataset to the actual performance of the full 30 minute dataset. Clearly, the 10 minute dataset can be used to predict the performance at longer test times (up to three times the available data in this example).

Table 30: 15 and 30 minute failure/survivor rates based on 10 minute data as compared to actual performance of the complete 30 minute dataset.

Test Time [Minutes]	Actual Failure Rate [% of Total Tested]	Predicted Failure Rate [% of Total Tested]
15	14.0%	14.0%
30	17.2%	17.2%

Based on the results shown in Table 30 the failure/survivor rate predictions appear valid as the error for these predictions is zero. This is not unexpected since, as mentioned above, the data appear to follow a single Weibull distribution. It is likely valid then to use the full 30 minute test data to predict the failure/survivor rates for longer test times. This will help to address the issue of test time described in Section 7.4.1. Specifically, it enables one to quantify the risk of using the shorter test time. Figure 73 shows the predictions for 45, 60, 75, and 90 minutes, using the full 30 minute dataset. These results are summarized in Table 31.

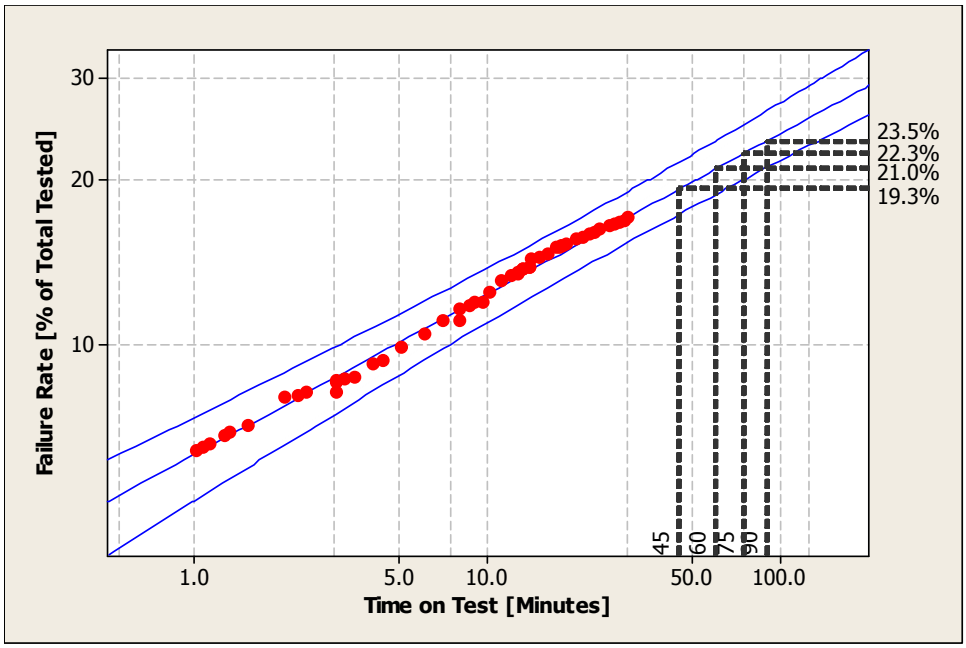


Figure 73: Probability plot of 30 minute time to failure data with extrapolation to 45, 60, 75, and 90.

Table 31: Summary of predictions for 45, 60, 75, and 90 minutes, and the difference in failure rate for each 15 minute increase.

Test Time [Minutes]	Predicted Failure Rate [% of Total Tested]	Delta Failures
30	17.2%	--
45	19.3%	2.1%
60	21.0%	1.7%
75	22.3%	1.3%
90	23.5%	1.2%

As the data in Table 31 show, a "law of diminishing return" effect applies to withstand data. By increasing the test time from 30 to 90 minutes the utility could expect to find an additional six failures for every 100 tested components. Such information allows the utility to weigh the benefit in terms of avoiding additional service failures versus the extra time needed to find them.

7.5 CLASSIFICATION

The process of classification may be approached a number of ways, but the final objective is the same: to assign every component to its correct group. The process involves three primary tasks:

1. Define the different subgroups into which the population of components will be classified.
2. Define rules to base the classification on.
3. Develop a procedure for evaluating components based on the set of rules.

In the case described here, the role of classification is to assess the condition of a component based on its physical characteristics. From this perspective, the first task is straightforward as the primary interest is in whether or not a component is about to fail. Therefore, one may use the typical "good" and "bad" groups. Before the remaining two tasks can begin, a set of data known as the training set is needed. This data must include measurements made on components whose true group membership is known. In the case of "good" and "bad", the training set must include a condition assessment of the components and the resulting service performance. In other words, the training set must answer two questions:

1. What was measured?
2. What happened to the component afterwards?

With such a training set in hand, it is possible to develop the rules and the procedure for evaluating those rules.

7.5.1 Classification Rules

The rules for classification can be based on any physical characteristic of the component. These characteristics may be evaluated using different diagnostic procedures or through different measurement data (features) obtained from a single procedure. In either case, different features may be used separately or together in order to define the

boundaries between classification groups. Figure 74 shows how multi-feature classification may be approached considering two features.

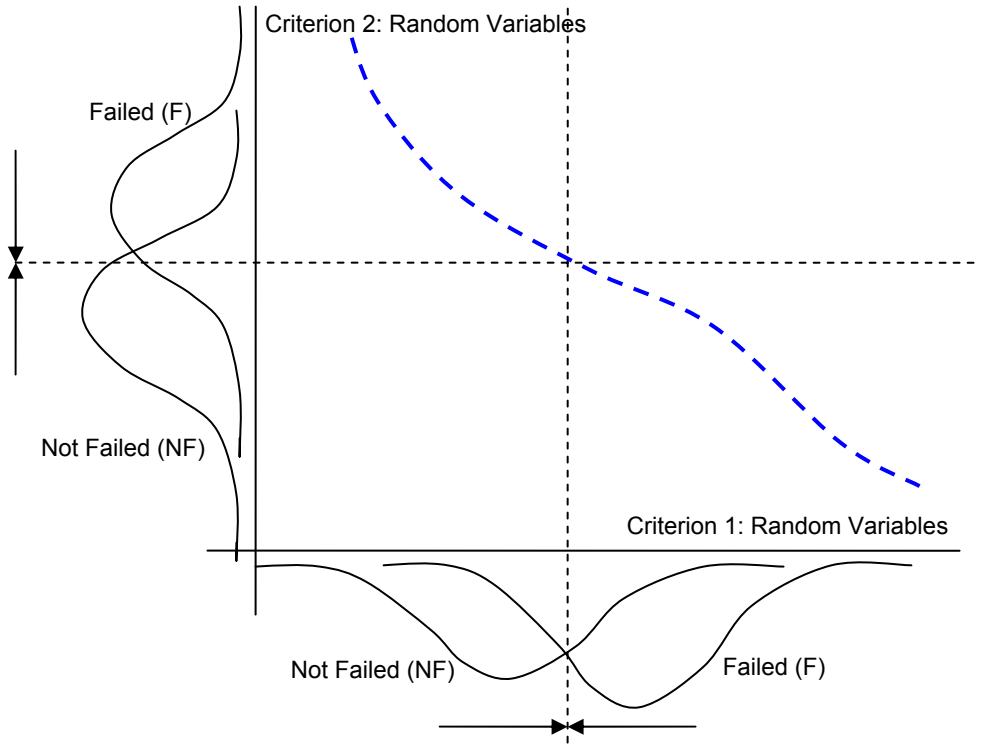


Figure 74: Example of multi-feature classification using two uncorrelated features with distributions for the two classification groups.

Note in Figure 74 that using one feature or the other alone may not provide sufficiently high confidence in the classification. However, combining the two can increase this confidence, especially if the new features represent diagnostic measurements taken using different techniques (in other words they are uncorrelated). In theory, the more features that can be used the more effective the classifier and, hence, the classification can be.

7.5.2 Classification Procedures

A number of procedures are available for classification including Bayesian, nearest neighbor, and Heuristic classifiers. These procedures either utilize the statistical characteristics of the data or other hidden properties that are identified through heuristic procedures such as self-organizing maps and neural networks. Regardless of the procedure, the classifier's goal is to define the boundary (illustrated in Figure 74) between classification groups that will enable the classification of a new data point that possesses a measurement for each feature.

7.5.2.1 Nearest Neighbor Example

A nearest neighbor classifier has been implemented in order to classify partial discharge measurements from underground cables. The nearest neighbors (or *k*-NN) method is a nonparametric method that classifies a data point as belonging to one group or another based on its distance from other samples whose group memberships are known. The basic procedure requires identifying the *k* (an odd integer) nearest samples to the data point that is being classified. In the classical *k*-NN algorithm, once these *k* samples are determined it is only necessary to determine which set the majority of the *k* samples belong to. The new data point is then classified as belonging to this same set.

Equation (7.8) describes the population of neighbors that will be used to classify a new data point, *X*, into one of *n* classes.

$$K = \sum_{i=1}^{n} K_i, \tag{7.8}$$

where,

K = Total number of neighbors under consideration,

K_i = Number of neighbors that belong to class *i,*

n = Total number of classes.

The classification process is defined as:

$$\text{Classification}(X) = \left\{ i \middle| \max_{i \in \{1,2,\ldots,n\}} K_i \right\}. \tag{7.9}$$

This example represents a two-feature type classification as two measurements are available: (1) charge magnitude (pC) and (2) inception voltage (p.u.). In this case, the actual classification for all samples is known. Therefore, it is possible to test this classifier using a subset of data as the training set and the remaining data as the testing set. The success rates for the two groups, "fail" ("bad") and "no fail" ("good"), are shown in Figure 75 for different numbers of neighbors.

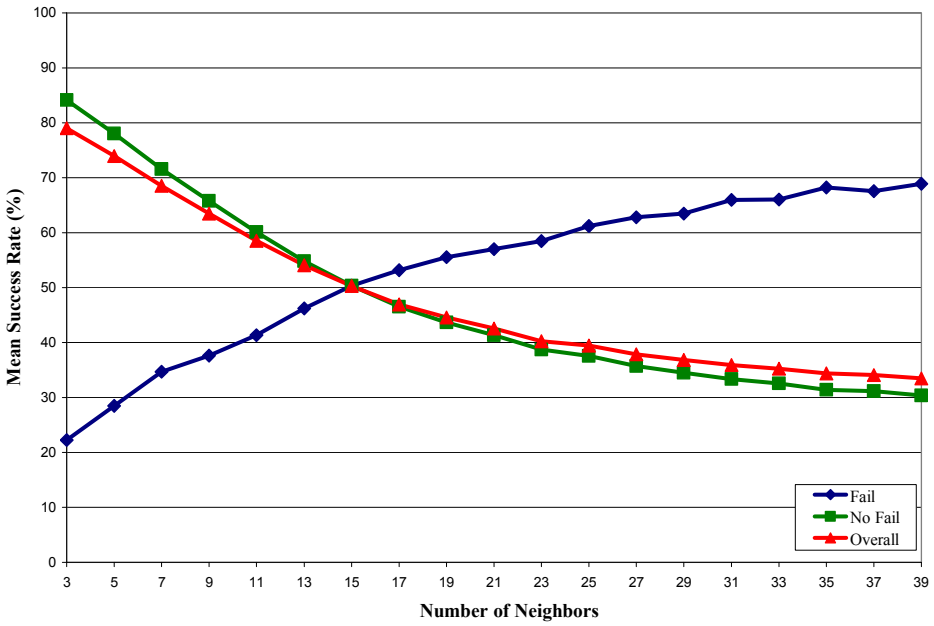


Figure 75: Sample two-feature classification success using two partial discharge features.

For this example, a balance in success rates for the two groups is achieved for 15 neighbors and a resulting success rate of approximately 50%. This is equivalent to an overall diagnostic accuracy of 50%. However, it is important to note that the two groups respond in opposite ways to changes in the number of neighbors. This is due to the

substantial difference in population size. The “no fail” group is approximately 10 times larger than the “fail” group. Differences of this nature are quite common in underground cables. With this mix, high overall diagnostic accuracy can be achieved be classifying all data as “good” but this provides little help for reliability. Clearly the two features used in this example do not contain enough information to enable high condition-specific diagnostic accuracies for both groups.

7.6 SUMMARY

This chapter has described several techniques for assessing different aspects of diagnostic test results. The basic concept is to relate diagnostic data to service performance after the testing is complete. In addition, some of the techniques may be used to aid in establishing suitable guidelines that can be included in industry standards. The techniques may be used with virtually any form of diagnostic test and component although they have been demonstrated using field data for underground cable systems.

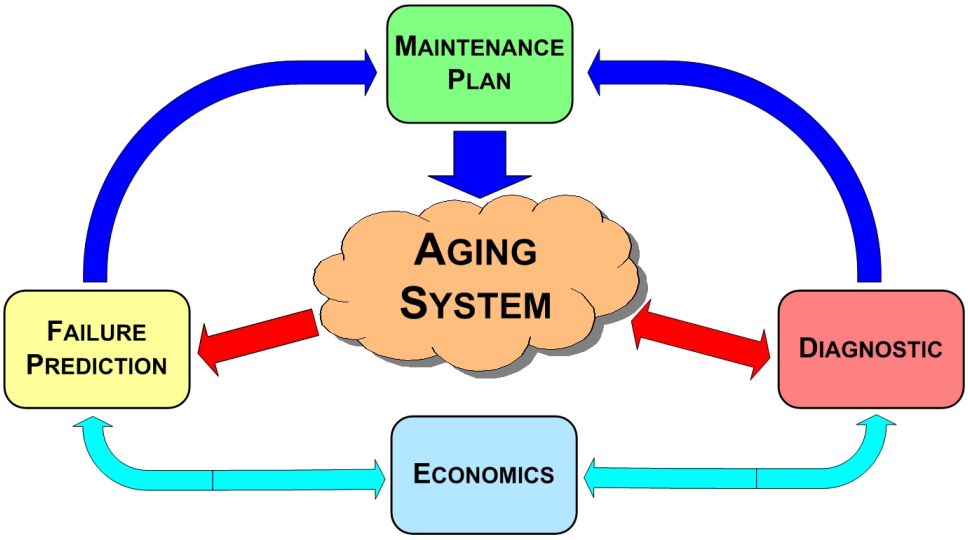
PART II SUMMARY

Part II has described the two primary functions of the diagnostic facet:

- Calculation of diagnostic accuracies.
- Interpretation of diagnostic field data and service performance data.

Together, these functions allow the diagnostic facet to determine the effectiveness of a particular diagnostic technique. This can be done using assessment techniques such as performance ranking, diagnostic outcome mapping, Weibull analysis, survivor analysis, and classifiers. The resulting diagnostic accuracies are vital to the operation of the economics facet that will be discussed in Part III. The use of diagnostic techniques in the economics facet is heavily dependent on the diagnostic technique's ability to correctly assess each component.

PART III: ECONOMICS FACET



The economics facet represents the primary decision and evaluation mechanism in the asset management chain. It relies on input from both the failure prediction and diagnostic facets to decide how best to construct the final maintenance plan. In addition, this facet sends feedback to both the other facets as it acts as the intermediary between them.

Understanding this process requires an explanation of the kind of situation a typical utility might face as it decides to embark on a program to improve reliability using diagnostics. By definition, the reliability of a component is defined as the probability that it will adequately perform its specified purpose for a specified period of time under specified conditions [11]. The goal of a reliability program is to maximize this probability over the chosen time period. As reliability itself and the process of repairing the system have associated costs, a second goal of the reliability program is to reduce the operating costs. As is described in Chapter 9, a service failure costs a utility much more than the time and materials needed to replace a failed component.

One approach to improving reliability is to proactively replace large areas of components with new components. This is the simplest reliability program to understand but it is also by far the most expensive. An alternative approach is to employ diagnostic tests to help select the components that are likely to produce failures in a chosen time horizon (generally several years). These programs have the potential to produce the same improvement in reliability (during the target time horizon) while minimizing the amount spent on dealing with those components that are likely to fail.

As is discussed in Chapter 10, diagnostic programs can generate economic savings as compared to both replacing entire populations of components and simple repair on failure. Unfortunately, there are situations where diagnostics can generate a loss. The choice of component population, or at-risk population, and diagnostic test are vital factors in whether or not a diagnostic program generates a savings. Chapter 9 describes how the

total cost of the diagnostic program may be computed and Chapter 10 illustrates several situations in which both savings and losses may be generated by the diagnostic program.

The following chapters will describe the development of a diagnostic program from the ground up, the necessary economic modeling, and illustrate both using case studies:

- **Chapter 8: The Diagnostic Program** – This chapter describes the development of a diagnostic program including the four phases of the program.
- **Chapter 9: The Economics of Diagnostic Programs** – This chapter introduces the cost functions and, more importantly, the benefit function used to calculate the savings/losses of diagnostic programs.
- **Chapter 10: Economic Simulation Studies** – The focus of this chapter is the demonstration of the benefit function using several case studies. Specifically, this chapter will demonstrate under what conditions diagnostic programs can generate savings and losses.

CHAPTER 8: THE DIAGNOSTIC PROGRAM

The focus of this chapter is the development of the stages of a diagnostic directed reliability strategy. This model allows a utility to develop a complete plan for executing their program. Furthermore, this program also forms the foundation for the calculation of the economic costs as well as the potential savings that a diagnostic program can provide. The economics aspects of the proposed approach shall be the topic of Chapter 9.

8.1 GENERAL DIAGNOSTIC PROGRAMS

Before explaining the phases of diagnostic programs, it is useful to examine how diagnostic programs function in general. Under ideal conditions, a utility would wish to employ a diagnostic test to help determine how best to proactively perform maintenance on its system. There are often other political reasons to do so, but this is beyond the scope of this discussion. The reasons behind this decision (cost of reliability) can be the result of penalties from the local regulator, industrial customer complaints, politics, media attention, or failure predictions. The utility has the primary goal increasing the ratio of reliability improvement to the number of maintenance or replacement actions employed.

The process begins with the identification of a subpopulation of components from the total population that the utility considers, based on the available data and engineering judgment, to be "at-risk" for producing failures in the near future. An example of such a population is shown in Figure 76 with the true condition of each component identified symbolically. Note that the utility does not know in advance of the program the true conditions of the components in its system.

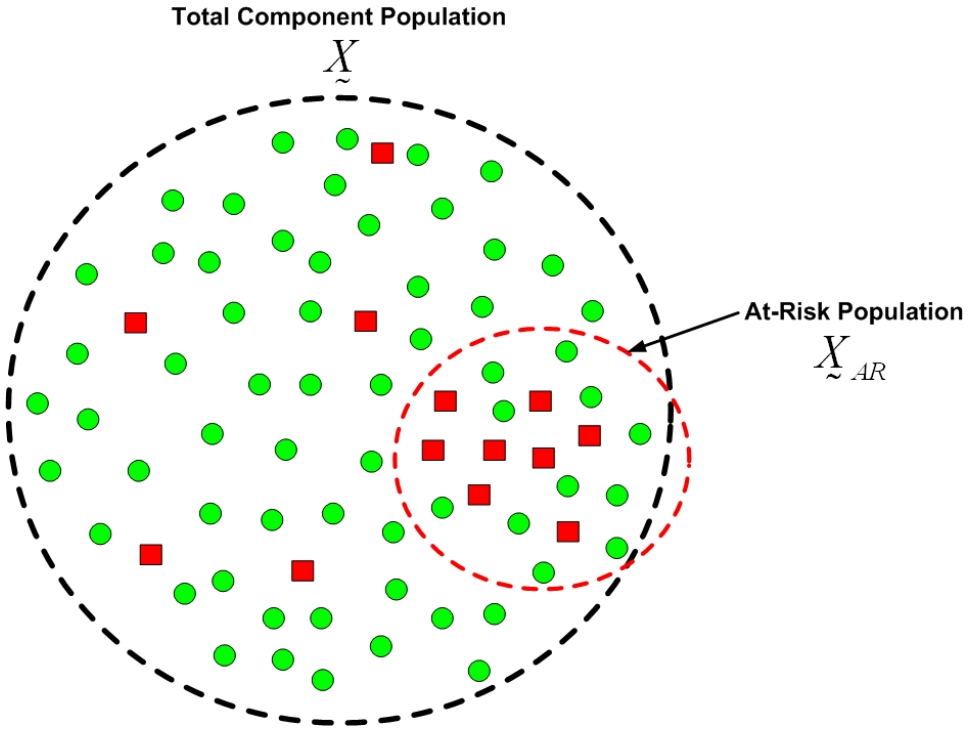


Figure 76: Sample component population with at-risk population identified via historical records. Components that do not require action are shown as dots (●) while components that do need action are shown as squares (■).

Note that the $\underset{\sim}{X}$ notation denotes a set of components.

The term "at-risk" in this case refers specifically to higher than average failure rates. In other words, the at-risk population designated in Figure 76 would have a local failure rate that is higher than the failure rate for the entire population. In their definitions of "at-risk," utilities may include other factors that cannot be modeled mathematically but can be termed as "human factors". This definition differs from the economic term "value at risk" in that "at-risk" in this context refers to the potential for failure of a component or group of components as opposed to the maximum financial loss for a given level of confidence in planning uncertain investments.

The next step in the program is to conduct diagnostic test(s) on the entire at-risk population. The results of the diagnostic testing are used to partition the at-risk

population, $\underset{\sim}{X}_{AR}$, into mutually exclusive sets into k classes whose possible conditions that may identified by the diagnostic. Each condition can be, loosely, associated with k different risks of failure (or hazard, defined as the probability of a component failing in the next time interval (Δt) given that it has survived to time t). Logically, each of the k conditions requires a different level of maintenance to improve the condition of the components in that condition enough to be able to survive, without failure, the desired number of years.

Without loss of generality, level of risk for each condition can be ranked such that set $\underset{\sim}{X}_1$ would contain the most degraded components while $\underset{\sim}{X}_k$ would include components that are like new. This can be stated mathematically as:

$$\begin{gathered}\text{Condition}\left(\underset{\sim}{X}_i\right) < \text{Condition}\left(\underset{\sim}{X}_j\right),\ j > i, \\ i \in \{1,2,\ldots,k-1\},\ j \in \{2,3,\ldots,k\},\end{gathered} \tag{8.1}$$

where,

Condition ( ) = Relative measure of how close the component is to being like a "new" component.

Figure 77 shows the cumulative probability of failure as a function of time for the k different conditions. Note that the components that are most severely degraded have the highest probability of failing in a short time. In this example, the probability curves are assumed to consist of only one failure mode, hence they are linear. In general, these probability curves may consist of multiple gradients and bends.

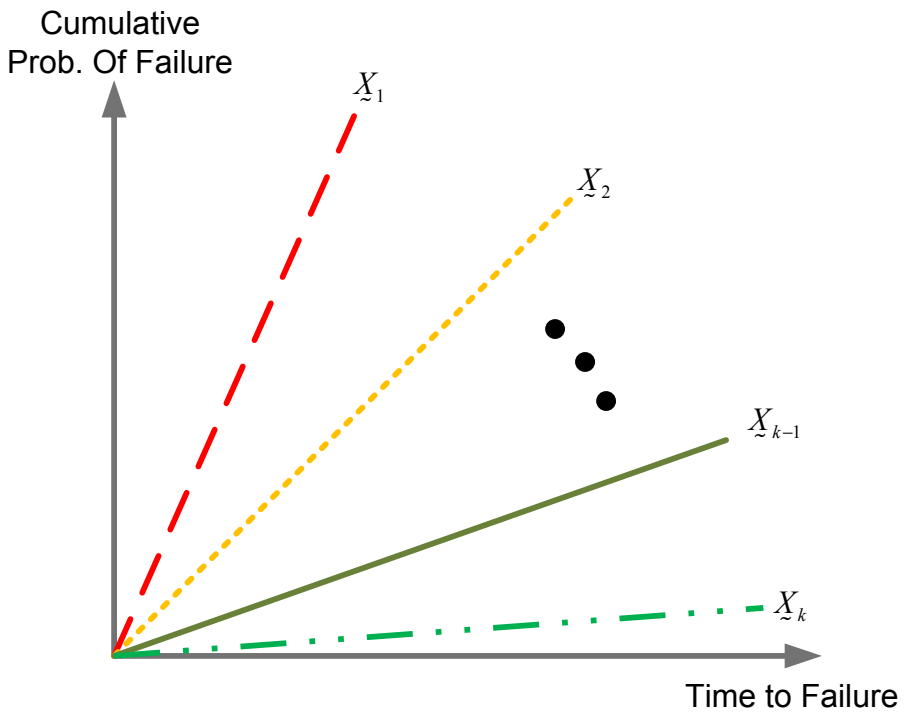


Figure 77: Probability curves for different subsets of the at-risk population. Set $\underset{\sim}{X}_1$ has the highest probability of failure at any given time.

The exact number of identifiable conditions (k), and hence sets, depends on the chosen diagnostic and the number of possible remedial actions that may be performed. The case of $k = 2$ is the simplest situation since the components are either classified as "bad" and replaced or as "good" and left unchanged.

The basic premise behind the use of the diagnostic is that relatively few components in the at-risk population will actually be classified into the $\underset{\sim}{X}_1$ subset and replaced. This means that the utility will perform far less costly corrective actions on the bulk of the population as compared to replacing the entire population. So by applying the (presumably inexpensive) diagnostic, the utility may determine the minimum level of action each component needs in order to operate reliably for the target time horizon. Under these conditions, the utility would experience a net savings over complete and immediate replacement of the entire at-risk population. Simultaneously, the utility would

see an identical improvement in reliability as the same number of future service failures would be avoided, in theory.

Implementing an effective diagnostic program requires careful planning as the potential economic benefits can easily lead to economic losses. The remainder of this chapter addresses the four stages of a diagnostic program:

- Selection – Choose the components whose reliability is questionable using all available historical data.
- Action – Decide what types of maintenance actions will be performed based on the system topology and prevalent failure mechanisms.
- Generation – Perform diagnostic tests to generate the data that will be used to determine what action will be performed on each component.
- Evaluation – Determine the effectiveness of the Selection, Action, and Generation. Does the program meet expectations in terms of reliability? Can the program be improved?

When combined these phases form the acronym, SAGE. Figure 78 shows the effect each of these phases on the failure rate for the at-risk population.

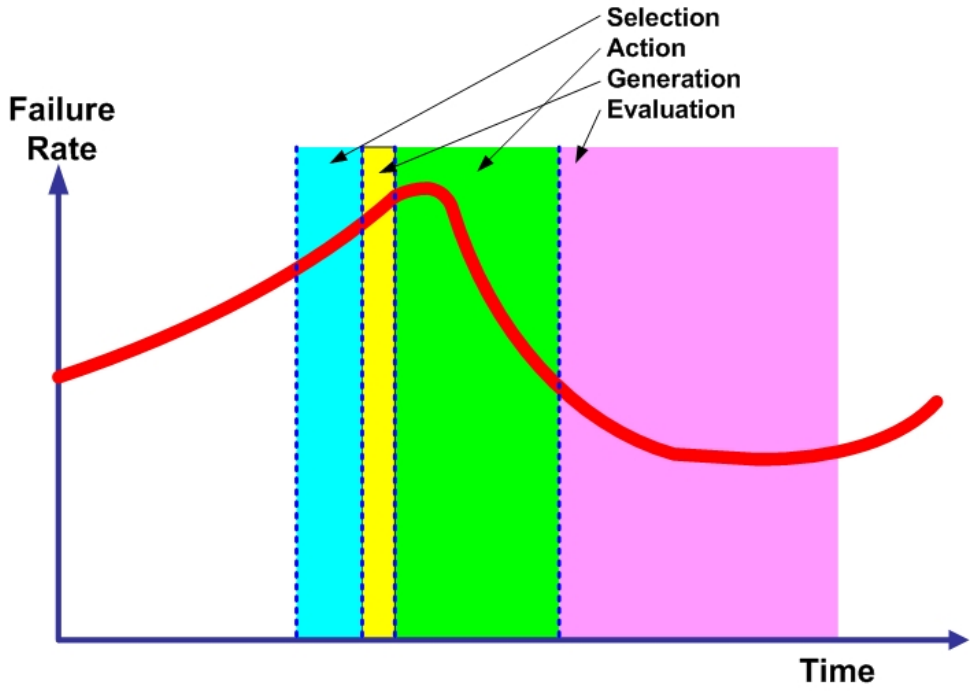


Figure 78: Effect of SAGE on the failure rate of the at-risk population.

Note that the failure rate in Figure 78 continues to increase during the selection and generation phases and then only after the actions are completed does the failure rate start to decrease. Furthermore, after some time the failure rate will begin to increase again and this will be observed during the evaluation phase.

These phases are discussed in the following sections.

8.2 SELECTION STAGE

The selection phase represents the initial attempt by the utility to identify those components in the network that are susceptible to failure within a chosen time horizon. The size and composition of this population greatly affect the possible improvements in reliability and the economic savings that could result from the program. The chosen subset of components represents the at-risk population, $\underset{\sim}{X}_{AR}$, defined earlier, that will be tested using one or more diagnostic tests.

In theory, selection begins with an assessment of the information already available within the utility and should address each of the following topics:

- System Construction – How are the components used in the system? Are components installed in parallel, series, or both? What level of redundancy is present? Can components be easily isolated from the network without impacting customers?
- Acceptability of Failure During Testing – Is a failure tolerable during testing if components are subjected to elevated test voltages?
- Available Historical Data – Number of components of the same type in service, their ages, and failure histories.
- Failure Projections – How fast are failure rates increasing? If there are multiple designs for the same component, in which are the failure rates increasing fastest?
- Prevalent Failure Mechanism – What causes the majority of failures? Is the mechanism electrical, mechanical, or thermal, in nature?
- Objective – Is the objective to improve reliability, reduce costs, or both? Has a budget already been allocated for this purpose?

Note that the above list does not consider human factors that would likely occur in a real utility. For obvious reasons, such issues are difficult to model and so will be excluded. Nevertheless, it is important to understand that these issues exist and they can exert tremendous influence over this process. For example, many network operators will be hesitant to allow work within their regions if that work has the possibility of producing a failure. Therefore, aversion to risk may be a driving force in the selection process. Another example that can occur is the inclusion of areas that may not actually be at-risk from a failure rate perspective, yet they contain customers that the utility does not wish to risk a service failure with. Therefore, regardless of the condition of the areas containing such customers, the utility will keep these regions as part of the program.

8.2.1 Prioritization

From a purely theoretical standpoint, the selection process should be focused on components with higher than average failure rates as these will yield the greatest improvements in reliability if they are acted upon. However, as mentioned above, the importance of a group of components is also a consideration. Figure 79 shows the local failure rates for different areas of a utility for a particular component type. According to this figure, the area labeled as χ has historically experienced the highest failure rate.

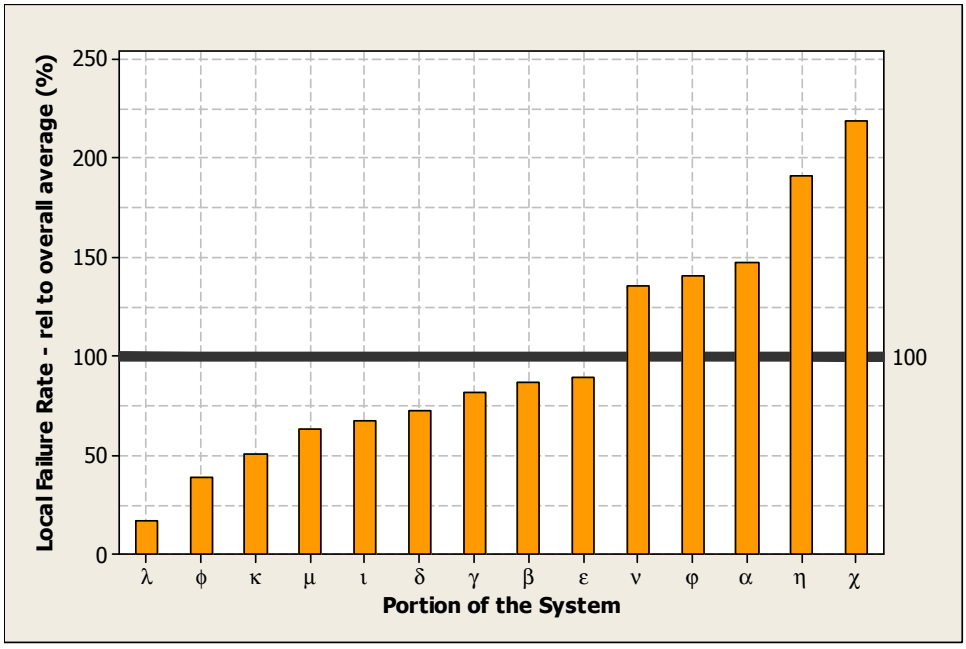


Figure 79: Example of local failure rates for different subsets of the total component population.

Figure 80 shows the comparison of each area with regards to both failure rate and importance.

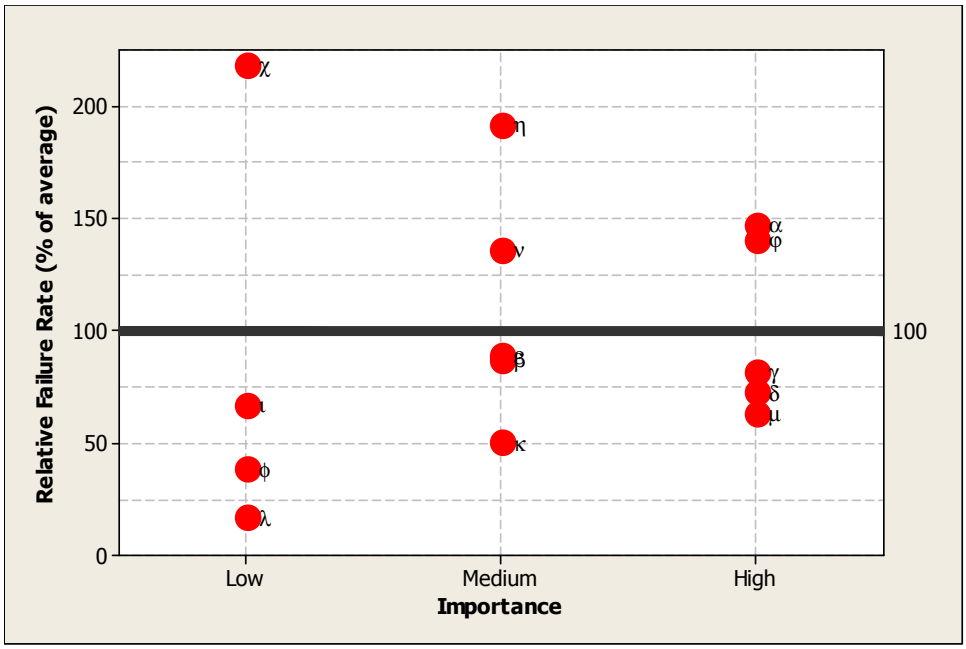


Figure 80: Comparison of local failure rates with the relative importance of each partitioned area to the utility.

In Figure 80, the highest failure rate area is of low importance to the utility. Therefore, the utility would be less inclined to include it as part of the diagnostic program. On the other hand, there are several highly important areas with above average failure rates that might be addressed in the diagnostic program before any other.

The identification of these components relies heavily on the availability of historical records and sound engineering judgment. It is well known that most utilities have not been inclined to maintain as detailed failure records as they should. This has changed only in the last few years. However, an engineer or regional operator with several years of experience within that particular utility can guide this process in the absence of suitable records. In addition, other criteria may be chosen in addition to local failure rates. These can include component age, design, operating stresses, or any other criteria (based on engineering judgment) that would adversely affect a component's reliability. If sufficient information exists within the records then the utility may utilize failure

projections. These may be obtained from the failure prediction facet described in Part I or Crow-AMSAA described in Section 7.2. These can aid the utility in defining where the system is heading in terms of failure rates as well as the amount of action (M_{AL}) that may be required to curtail an unacceptably high failure rate.

8.2.2 Subdividing the Population

To determine the at-risk population (and any other subpopulation) it is necessary to return to set theory. Define the set $\underset{\sim}{X}$ as the set of all installed components currently operating in the system as:

$$\underset{\sim}{X} = \left\{ x_1, x_2, x_3, ..., x_n \,|\, x_i = \begin{cases} 0 & x_i \text{ is unlikely to fail} \\ 1 & x_i \text{ is likely to fail according to a specified criterion} \end{cases} \right\}, \qquad \textbf{(8.2)}$$

where,

n = total number of components in the system.

Set $\underset{\sim}{X}$ can be partitioned according to the criteria described above to identify two main categories:

(i) Components that are at-risk, i.e. the historical records or engineering judgment indicate that the component is prone to fail (set $\underset{\sim}{X}_{AR}$ in Figure 76).

(ii) Components that are <u>not</u> under risk, i.e. it is expected that the component will continue operating without failing.

Then it is possible to define the partition of $\underset{\sim}{X}$ as:

$$\underset{\sim}{X}_{AR} = \left\{ x_j : j \in \underset{\sim}{J}_1 \right\},$$

$$\underset{\sim}{X}_0 = \left\{ x_j : j \in \underset{\sim}{J}_2 \right\}.$$

Moreover:

$$x_i = \begin{Bmatrix} 1 & \text{if the component is prone to fail} \\ 0 & \text{otherwise} \end{Bmatrix}.$$

The partition of $\underset{\sim}{X}$ then satisfies:

$$\underset{\sim}{X}_{AR} \cap \underset{\sim}{X}_0 = \varnothing,$$

$$\underset{\sim}{X}_{AR} \cup \underset{\sim}{X}_0 = \underset{\sim}{X}.$$

Based on these definitions, parameter ρ may be defined as:

$$\rho = \frac{\left|\underset{\sim}{X}_{AR}\right|}{\left|\underset{\sim}{X}_{AR}\right| + \left|\underset{\sim}{X}_0\right|} = \frac{\left|\underset{\sim}{X}_{AR}\right|}{\left|\underset{\sim}{X}\right|}. \tag{8.3}$$

As a result, ρ constitutes the fraction of the total population of X (where $X = \left|\underset{\sim}{X}\right|$) components that are considered to be at-risk. Even though the selection of this subset of X_{AR} (where $X_{AR} = \left|\underset{\sim}{X}_{AR}\right|$) components is done using the best available information within the utility, it is expected that this population will contain components that will fail and others that will not. The job of the diagnostic test will be to separate them so the utility may repair or replace the degraded components.

It is also important to carefully consider the size of the at-risk population. Larger populations will require more time for testing and possibly longer delays between testing and completion of the required maintenance and replacement actions. This time delay can be long enough to allow for additional service failures to occur either before the component is tested or between the time testing is completed and the completion of the corrective action. On the other hand, too small an at-risk population may not be economically justifiable if there are either too few "good" components or too few "bad" components. One may define the "good to bad" ratio, or G/B, as the percentage of "good" components as compared to the percentage of "bad" components in the population. In this case, the term "bad" refers to any component that requires either a maintenance or

replacement action. A G/B ratio of 10/90 leads to low benefit (if any) as the cost of diagnostic testing is simply added to the cost of replacement of virtually the entire population. On the other hand, the case of G/B ratio equal to 95/5 produces only small improvements in system reliability so the utility would need to look elsewhere to achieve its reliability goal.

There is no universal G/B ratio that can be applied to any diagnostic program. The actual range of suitable G/B ratios depends on the cost elements that will be discussed in Chapter 9. Chapter 10, on the other hand, demonstrates several cases where suitable G/B ratios can be observed. However, based on experiments and studies of diagnostic programs currently operating in US utilities a system that is at best 85/15 is suitable for inclusion in a diagnostic program. Unfortunately, the diagnostic accuracy must be in the range of 95% for systems with higher G/B ratios to yield economic benefit.

Once the selection process is complete, the utility must then examine what will be done to correct degraded components identified by the diagnostic test. This corresponds to the action phase.

8.3 ACTION STAGE

The action stage of the SAGE process refers to the establishment of the set of possible maintenance and/or replacement actions ($\underset{\sim}{M}$) that may be performed based on the results of diagnostic testing. Ideally, a specific action will be selected for each possible component condition. This defines the set $\underset{\sim}{M}$ as:

$$\underset{\sim}{M} = \left\{ M_i,\ i \in \{1, 2, 3, ..., k\} \right\}, \tag{8.4}$$

where,

M_i = Action that will be performed on components with diagnosis i.

Following the logic of ranking subgroups of the at-risk population based on condition the M_1 action would correspond to replacement of the component. In general, each corrective action carries with it a unique cost to complete ($C_{m,i}$) and an associated level of reliability improvement (ΔF_i). These two factors are, in theory, correlated to each other.

The number of actions is chosen based on the type of component, system design, and maintenance policies (or traditions) of the utility. The first and most influential of these issues is the type of component under investigation. All components, no matter how large or expensive, can be replaced given enough financial backing. However, many components exist in which a repair either is not possible or prohibitively expensive (i.e. as much as a new component) such as fuses or direct buried cable. Furthermore, the complexity and cost of the device will dictate whether or not multiple levels of repair should be employed. For example, three phase network transformers may require a repair as trivial as new oil or bushings or may receive more extensive repairs such as the rewinding of one phase. No matter the level of repair, a repair should never exceed the cost of replacing the entire component. Ideally, a component would receive the minimal amount of action needed to ensure reliable operation for a specified period. Unfortunately, in practice the repair techniques may not be as reliable as a new component and would correspond to the infant mortality region of the Weibull bathtub curve described in Section 7.3.

The system design will also dictate whether or not repairs are logistically feasible. For critical components the repair time may be too long or too difficult to complete onsite and, therefore, would require a new or previously refurbished device. The component may be so critical that only during certain times of the day or year can it be switched out from service to be repaired. This is certainly a suboptimal policy from a maintenance standpoint. However, it is a necessary tradeoff between maintainability and real-time operation.

The last issue, utility policies, will also exert influence over the possible maintenance actions. Maintenance crews may be trained to perform only certain repairs or the utility may have chosen to abandon certain component designs in favor of more advanced ones. There are a number of reasons for a utility to choose not to repair or to have specific levels of repair they are willing to perform. Many of these are due to the human factors mentioned earlier.

Once a suitable maintenance policy for each diagnosis has been established, work may be performed in the at-risk population to generate the diagnostic data. This constitutes the generation phase.

8.4 GENERATION STAGE

The generation stage of the SAGE process represents the point at which a suitable diagnostic is chosen and then performed on the at-risk population of components. By definition, the diagnostic technologies measure specific characteristics of the component that are believed to be symptomatic of the known failure mechanisms. These symptoms can generally be classified into two categories: (1) distributed and (2) local. Distributed symptoms are those that cannot be localized to specific sections of the component. Dielectric loss is an example of a global characteristic for underground cables. On the other hand, local symptoms can be attributed to specific portions of the component and the diagnostic is able to identify the portion of the component that is causing the symptom to appear. In underground cables, partial discharge is an example of a local condition.

The following factors should be considered during the generation phase:

- Prevalent Failure Mechanism: Global or local defect? Can the diagnostic measure a characteristic of the component from which its condition may be reliability ascertained?

- Accuracy of the Diagnostic: How often does the diagnostic correctly classify the component's condition?
- Cost of the Diagnostic: Does the cost of the diagnostic represent a large portion of the replacement cost of the component?
- Resolution of the Diagnostic: Does the diagnostic provide enough information to classify the components into the number of desired subpopulations?
- Reliability: Can the diagnostic be employed in the field to produce useful results?
- Risk: What is the risk of failing the component while on test?

An example of selecting a diagnostic test is shown in Table 32 for underground cable with different insulation materials.

Table 32: Diagnostic Tests for Single Cable Insulations [4].

	Withstand		**Partial Discharge**		**Dielectric**			
Insulation	AC 60 Hz and VLF	DC	Offline	Online	Dissipation Factor	Recovery Voltage	Polarization Current	DC Leakage
1	√	√	√	√	√	√		√
2	√		√	√	√		√	
3	√		√	√	√			
4	√		√	√	√			

From this table, it is clear that not all diagnostics are applicable to all insulations. As an example, consider the DC Withstand diagnostic in Table 32. If one's system were composed of cables insulated with insulation one then the DC Withstand test is reasonable to use. However, if the system under investigation used insulation two instead,

then the DC Withstand is an inappropriate diagnostic because of its adverse effect on the insulation material [4].

The above list of issues can be summarized as follows: Is the diagnostic able to diagnose the prevalent problem in the components and do so with high enough accuracy to provide an advantage to the program? The accuracy of the diagnostic is a critical factor. It is important, therefore, to consider the situation where the diagnostic is imperfect with regards to identifying "good", "bad", or both. This translates to acting on components that do not require it, but more importantly, leads to not acting where it is needed. Depending on the goals of the utility, one situation may be more tolerable than the other. Regardless of the preference, the asset manager would like to know the probabilities associated with performing a particular maintenance action and the chances that the actions are correct. Unfortunately, to compute these probabilities, the utility must have some knowledge about the accuracies of the diagnostic and the "true" numbers of components within each subpopulation. Suppose for the case of two possible conditions, "good" and "bad", the following variables are defined to represent the components within the at-risk population that correspond to those conditions.

$\underset{\sim}{X}_b = \{x_j \in \underset{\sim}{X}_{AR} : x_j \text{ will fail}\}$ = Set of components in the at-risk population that are "bad" and will fail in the near future.

$\underset{\sim}{X}_g = \{x_j \in \underset{\sim}{X}_{AR} : x_j \text{ will not fail}\}$ = Set of components in the at-risk population that are "good" and so will not fail in the near future.

$X_b = |\underset{\sim}{X}_b|$ = The number of "bad" components in the at-risk population.

$X_g = |\underset{\sim}{X}_g|$ = The number of "good" components in the at-risk population.

$X_b + X_g = X_{AR}$ = Each component is either "good" or "bad" but not both (i.e. $\underset{\sim}{X}_b \cap \underset{\sim}{X}_g = \varnothing$ and $\underset{\sim}{X}_b \cup \underset{\sim}{X}_g = \underset{\sim}{X}_{AR}$).

Probabilities can they be defined for the possible outcomes of the diagnostic test as shown in Table 33.

Table 33: Summary of probabilities for a 2-level diagnostic.

Conditional Probability	Definition	Description
p_g	$P(G \mid T < 0)$	Correct diagnosis of "good" component as good.
1-p_g	$P(G \mid T > 0)$	Incorrect diagnosis of "good" component as bad.
p_b	$P(B \mid T > 0)$	Correct diagnosis of "bad" component as bad.
1-p_b	$P(B \mid T < 0)$	Incorrect diagnosis of "bad" component as good.

In reality, the true values of the above probabilities will not be known prior to initiation of the program. However, based on other test programs one can establish the distribution of accuracies a particular diagnostic delivers. Note that, in general, the diagnostic may have different condition-specific accuracies as described in Chapter 6. Accuracy is vital to the selection of a diagnostic test as the maintenance decisions will be made based on the results of the diagnostic test. Using the diagnostic accuracy, it is possible to estimate the probability of performing each type of corrective action, as described in the following section.

8.4.1 Accuracy Revisited

Suppose the following set of maintenance actions will be performed on an at-risk population of components:

$$\underset{\sim}{M} = \{M_1, M_2\}, \tag{8.5}$$

where,

M_1 = Replace the component,

M_2 = Take no action.

The probability of performing a particular maintenance action also depends on the "true" composition of the at-risk population. The "true" composition is defined as the numbers of components belonging to each diagnosed subpopulation assuming the diagnosis was performed using a 100% accurate diagnostic. Combining the accuracies and population composition is done using Bayesian principles. It is necessary to make the assumption that the component survives the diagnostic test procedure. This probability may be written as:

$$P[\text{Performing } M_1] = \frac{(1-p_g)X_g + p_b X_b}{X_{AR}}, \tag{8.6}$$

where,

X_g / X_{AR} = Fraction of the at-risk population that is truly "good,"

X_b / X_{AR} = Fraction of the at-risk population that is truly "bad."

The same approach may be applied to the second maintenance action (M_2) to obtain:

$$P[\text{Performing } M_2] = \frac{p_g X_g + (1-p_b)X_b}{X_{AR}}, \tag{8.7}$$

For a two-action program the probabilities of the two possible maintenance actions are complementary.

$$P[\text{Performing } M_1] = 1 - P[\text{Performing } M_2] \tag{8.8}$$

For higher numbers of subpopulations, the process is the same, except that probabilities must be defined that relate all of the possible conditions to all possible misdiagnoses. In general, the utility will only have information on the overall diagnostic accuracy from other test programs. This information is sufficient to calculate the various probabilities related to the diagnostic itself. Using this information, it is possible to make predictions as to the relative sizes of each diagnosis group. This then allows one to estimate the costs associated with each group of corrective actions.

8.5 EVALUATION STAGE

The final stage of the SAGE process is the evaluation stage. This is the stage in which the utility engineers should ask themselves: Are we getting what we expected? A question such as this covers many issues; however, these can be summarized by two key topics: (1) Cost and (2) Reliability. A diagnostic program must deliver improved reliability at a lower cost as compared to other maintenance strategies in order to be considered effective. Evaluation tools such as those presented in Chapter 7 may be used to assess the impact the program has made on the system reliability. As a result, the diagnostic program represents an optimization problem of maximizing the reliability improvement and also a minimization problem in terms of the costs involved. Between these two it is possible to quantify the benefit the program has produced for the utility.

In addition to the economic benefits, the evaluation phase also consists of the continued monitoring of the tested components. Through this monitoring, additional information can be gathered on the diagnostic accuracy that would allow for more refined predictions in future programs. Also, this monitoring aspect allows the utility to gather data that may be used in conjunction with either failure prediction model to predict when the at-risk population will again require attention. This is a critical step since those components that did not require corrective actions will continue to age and will eventually require them. It is valuable to be able to identify the point at which a reassessment of these components should be carried out. The evaluation phase represents an ongoing process that must be carried out until the need again arises to conduct another diagnostic program.

8.6 SUMMARY

This chapter has described the four stages of a diagnostic program: (1) selection, (2) action, (3) generation, and (4) evaluation. Each stage is crucial for the successful implementation of a diagnostic program. This is demonstrated in Chapter 9 and Chapter

10 as the cost components and simulation studies are presented. The following chapter, Chapter 9, develops the specific cost elements that must be used in the calculation of economic benefit (or savings) for power system equipment.

CHAPTER 9: THE ECONOMICS OF DIAGNOSTIC PROGRAMS

This chapter introduces the cost model used in the economic modeling of diagnostic programs. The ultimate goal of this research is the calculation of economic benefit defined as financial savings resulting from a lower total cost as compared to an alternative program. Section 9.1 develops the mathematical definitions of the cost elements needed for the economic model. Section 9.2 addresses the issue of savings resulting from improved reliability while Section 9.3 and Section 9.4 examine the calculation and interpretation of benefits of using diagnostic techniques with respect to alternative asset management options..

9.1 DIAGNOSTIC PROGRAM COST COMPONENTS

This section describes the calculation of the cost components associated with the diagnostic program beginning with the cost of a service failure. It should be noted that each of the following cost components will contain some uncertainty about their true values at the time when asset management options are being evaluated and the potential benefits assessed.

9.1.1 Cost of Service Failure

The cost of a single service failure is the most difficult cost to compute as a significant portion of this cost it intangible. The intangible portion arises from the need to include a customer penalty cost for unreliability. In work performed by NEETRAC as part of the CDFI project, attempts have been made to quantify this element. However, utilities are unable to define a strict dollar amount for this cost. With this uncertainty in mind, the cost per failure can be expressed as:

$$C_F = C_{FR} + C_S + C_{Cust} \cdot N \qquad \textbf{(9.1)}$$

where,

C_F = Total cost per failure $[\$/\text{Failure}]$,

C_{FR} = Cost of repairing the component when it has failed $[\$/\text{Failure}]$,

C_S = Switching cost of outage $[\$/\text{Failure}]$,

C_{Cust} = Penalty resulting from customer relations issues $[\$ / \text{Customer} / \text{Failure}]$,

N = Number of customers impacted by the outage.

Equation (9.1) can be decomposed into two distinct parts as:

$$C_F = \underbrace{C_{FR} + C_S}_{\text{Cost of Restoring Service}} + \underbrace{C_{Cust} \cdot N}_{\text{Failure Consequence}}. \quad \textbf{(9.2)}$$

The first portion of (9.2) represents the cost of material and labor needed to repair the failure as part of the service restoration process. This cost includes both the man hours and materials needed to restore the failed component to operation again. It is a cost that would be incurred by the utility regardless of whether the defect was identified through diagnostic testing or the service failure.

On the other hand, the second set of terms in (9.2), the "failure consequence," represent additional losses incurred because the failure happened while the component was in service. These include the losses resulting from un-served load and emergency switching activities as well as penalties both from the local regulator and possibly from industrial customers. The penalty costs together are known to be significant with respect to the costs associated with restoring service and repairing the failed component. Unfortunately, some of this information is kept confidential by utilities and regulatory agencies. The regulator cost depends on many factors including past performance of the utility and current failure rates. These are measured through various reliability indices such as SAIFI, CAIDI, etc. [2], [27], [39].

9.1.2 Cost of Diagnostic Testing

The costs of diagnostic testing the entire at-risk population ($\underset{\sim}{X}_{AR}$) can be written as:

$$C_D = X_{AR} \cdot \left[C_T + C_S \right], \tag{9.3}$$

where,

C_D = Total cost of performing the diagnostic test on the at-risk population $\left[\$\right]$,

C_T = Cost of diagnostic equipment and personnel $\left[\$/\text{Test}\right]$ or $\left[\$/\text{Component}\right]$,

C_S = Cost of line crew for switching the circuit out of service, if needed $\left[\$/\text{Test}\right]$.

These costs are assumed to be known prior to the initiation of the diagnostic program.

It should be noted that the cost of the testing equipment and personnel can vary significantly between diagnostic techniques. Some techniques can be performed using small inexpensive equipment operated by utility crews while others require far more expensive equipment and technical expertise. In theory, the selection of a diagnostic following the latter case should be justified by either improved accuracy or higher suitability to the system and prevalent failure mechanism.

9.1.3 Cost of Corrective Actions

The costs associated with the multi-tiered approach to maintenance are described by:

$$C_M = \sum_{i=1}^{k-1} C_{M,i} \left| \underset{\sim}{X}_i \right|, \tag{9.4}$$

where,

C_M = Total cost of maintenance performed using multi-tiered approach,

$C_{M,i}$ = Cost of performing the required corrective action for components in condition i,

$\left|X_i\right|$ = Number of components in condition i.

Note that the costs shown in (9.4) only reflect the cost of performing a particular level of corrective action on all components diagnosed as requiring it. Also, the summation runs until the *k-1* subpopulation as the k^{th} subpopulation is defined as the set of components that do not require action. Therefore, the cost $C_{M,k}$ is identically zero. These costs are, again, known before the start of the diagnostic program.

9.1.4 Expected Cost of Maintenance and Testing

The cost of maintenance and testing were defined in (9.3) and (9.4), respectively. They are combined as:

$$\begin{aligned} C_{AR} &= C_D + C_M, \\ &= \underbrace{X_{AR}\left(C_T + C_S\right)}_{\text{Cost of Diagnostic Testing}} + \underbrace{\sum_{i=1}^{k-1} C_{M,i} X_i}_{\text{Cost of Maintenance}}, \end{aligned} \tag{9.5}$$

where,

C_{AR} = Cost of diagnostic program in at-risk population [$].

It should be noted that in (9.5) C_D is a fixed cost that will be incurred regardless of the results of the maintenance performed. For a two-action diagnostic program, the cost of maintenance, C_M, may be rewritten considering the condition-specific accuracies as:

$$\begin{aligned} C_M &= C_{M,b} X_{AR} \left(\frac{\left(1 - p_g\right) X_g + p_b X_b}{X_{AR}} \right) \\ &= C_{M,b} \left(\left(1 - p_g\right) X_g + p_b X_b \right), \end{aligned} \tag{9.6}$$

where,

C_M = Total cost of corrective actions performed during the diagnostic program,

$C_{M,b}$ = Cost of corrective action performed on components diagnosed as "bad,"

p_g = Condition-specific accuracies for components diagnosed as "good,"

p_b = Condition-specific accuracies for components diagnosed as "bad,"

X_g = Number of "good" components in the at-risk population,

X_b = Number of "bad" components in the at-risk population.

9.1.5 Total Cost

Using the cost components developed in the last four sections, it is possible to begin to construct a cost diagram that shows the accumulation of these costs and their relative sizes. Figure 81 shows the cost components from Section 9.1.4.

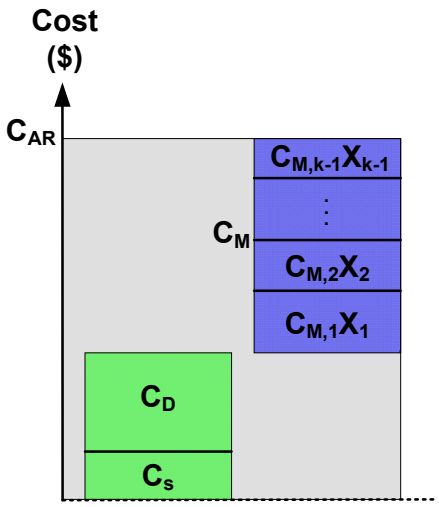


Figure 81: Summary of diagnostic program costs including C_D (diagnostic test), C_S (switching crew for diagnostic), C_M (maintenance), and C_{AR} (work on at-risk population).

Additional cost components related to reliability must be added to those shown in Figure 81 as they too will contribute to the total program cost. These components are described in the following sections.

9.2 RELIABILITY ISSUES

Reliability improvements are, in theory, quantified exclusively by changes in failure trends observed since the diagnostic program began. Techniques such as diagnostic outcome mapping (Section 7.2), Weibull analysis (Section 7.3), and failure prediction (Part I), can all be employed to assess these changes. In addition, industry reliability indices such as MAIFI, CAIDI, and SAIFI, will also reflect changes in system reliability.

The factors that must be considered in evaluating the reliability improvement that may or may not result from a diagnostic program are as follows:

- The maximum reliability improvement is limited by the number of components within the at-risk population that will fail in the target time horizon.
- If the at-risk population represents a subset of the total population then failures will still occur in components that were not tested.
- Not all components that receive corrective action will be sufficiently restored to survive through the entire target time horizon.
- Inaccuracy in the diagnostic test will lead to "bad" components being diagnosed as "good" and vice versa. These misdiagnosed "bad" components will produce failures in the at-risk population.

These issues are each discussed in the following sections.

9.2.1 Maximum Reliability Improvement

Within any population of components there exists a maximum level of improvement that the population can experience. This is directly related to the number of "bad" (or "not good") components in the at-risk population:

$$\text{Reliability Improvement} \propto \frac{X_b}{X_{AR}} \qquad 0 \le X_b \le X_{AR} \tag{9.7}$$

Equation (9.7) shows that the potential reliability improvement is directly proportional to the number of "bad" components that are present in the at-risk population. "Good" components do not adversely affect the reliability of the system. Therefore, acting on them does nothing to improve the reliability. As a result, the maximum reliability improvement is seen when the entire at-risk population is "bad." On the other hand, the minimum improvement is seen when there are no "bad" components, only "good" ones. Neither of these scenarios is truly beneficial for the utility. The latter case produces no improvement in reliability while the former case produces no benefit in terms of savings as will be discussed in Section 9.3.

From the reliability standpoint, the goal is to identify as many truly "bad" components in the system as possible and to perform only the minimum corrective action needed to guarantee their reliable operation during the target time horizon.

9.2.2 Failures Missed by Selection Phase

Unless the at-risk population includes all the components in the system the utility should expect failures to still occur. As mentioned earlier, during the selection phase of the program it is important to select the components that are at-risk of failure in the near future. Unfortunately, for any subpopulation the utility identifies, there is a chance that the components outside of the at-risk population will unexpectedly fail. This situation is illustrated in Figure 82.

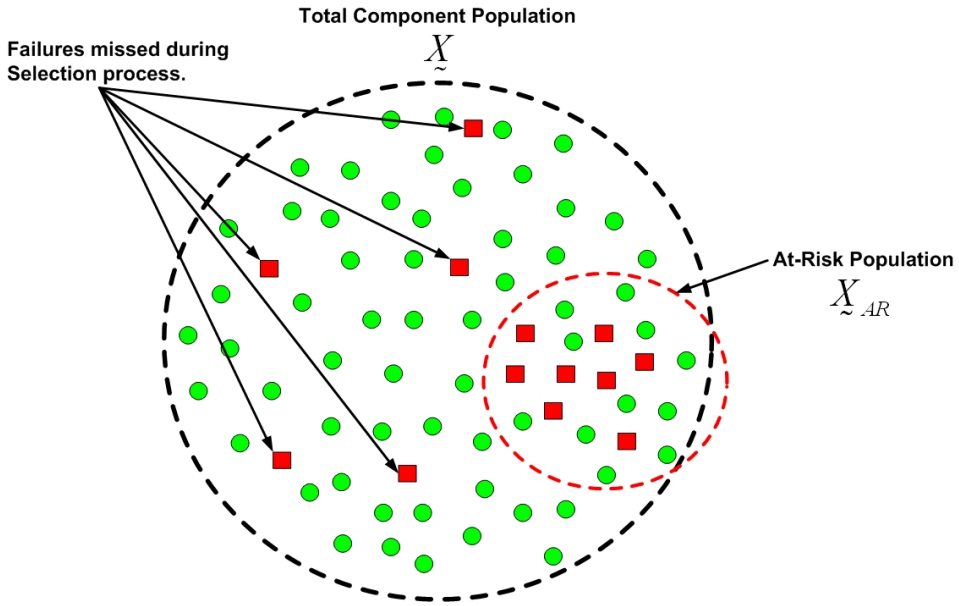


Figure 82: Sample component population with failures occurring on components that were not included in the at-risk population identified during the selection phase.

The scenario portrayed in Figure 82 is likely to occur as the records are never sufficiently detailed to allow for a perfect identification, hence the reason for employing diagnostics. On the other hand, the way to ensure that all the failures are included in the at-risk population is to consider the entire population as being at-risk for failure. However, given the population sizes that are typical for utility systems this becomes impractical. Therefore, the at-risk population will have, in theory, the highest priority for testing. Until this point, the testing of the at-risk population has been assumed to be possible in a relatively short time frame, perhaps one year or less, so that additional aging of the components during the testing program could be neglected. However, if the population is large enough, as with most components within a typical utility's system, the task of testing will likely take several years to complete with reasonable annual financial resources. The effect is that components that appeared to be unlikely to fail at the start of the testing program may, in fact, age enough during the test program to make them

susceptible to failure as discussed earlier. Therefore, the effect of the diagnostic program on the system-wide reliability will be reduced.

9.2.3 Failures Missed and Created During Action Phase

The chosen set of maintenance actions ($\underset{\sim}{M}$), the utility's ability to perform them correctly, and the manufacturing quality of the replacement components, will also impact the reliability yield. Power system devices (the components and the installation) follow the well known Weibull "bathtub" curve as shown in Figure 83.

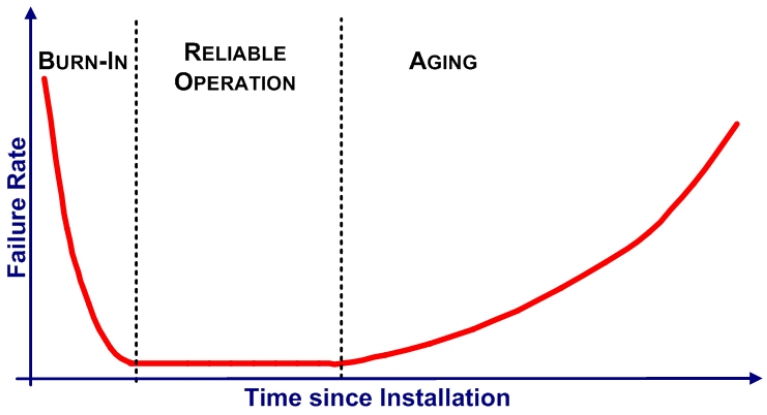


Figure 83: Illustrative Weibull "bathtub" curve showing failure rates for the various stages of component lifetime.

Using Figure 83, the goal of the diagnostic program can be defined as follows: identify the components that are farthest into the *aging* region and then perform the necessary corrective action to return them to the *reliable operation* region. Unfortunately, the "bathtub" (failure hazard) curve shows that new components can experience higher than normal failure rates (as discussed in Section 7.3) for a short period following installation. Failures during this stage are usually due to manufacturing or workmanship defects. Hence, performing a replacement on a component that is not far enough into the

aging region may, in fact, precipitate a failure sooner than it would have occurred had the old component remained in service. This is illustrated in Figure 84.

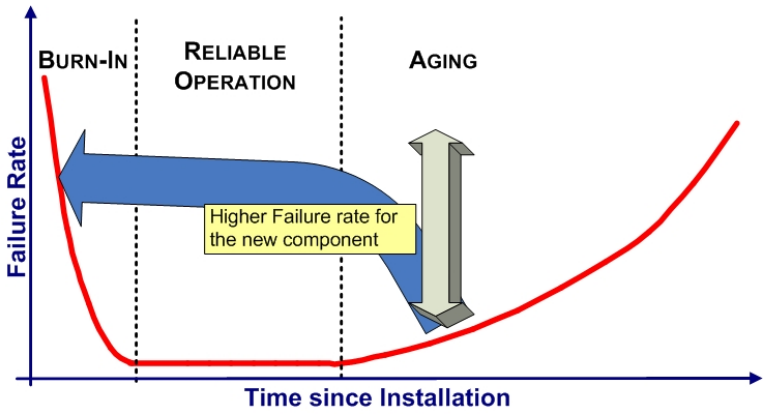


Figure 84: Graphical interpretation of replacing component in early portion of *aging* region with new component possessing high *burn-in* failure rate.

After a long period of *reliable operation* the components will reenter the *aging* stage where the failure rate will once more begin to increase. Given the behavior illustrated by the bathtub curve, the utility can expect to see a partial reduction in the expected number of avoided failures if it has decided to replace a large percentage of the at-risk population.

In addition to the failures that occur during "*infant mortality"* stage, there is also the possibility in diagnostic programs that employ more than two action levels that the chosen corrective action may not be aggressive enough to bring the component back to *reliable operation*. In this case, the component simply returns to an earlier point within the a*ging* region, as depicted in Figure 85.

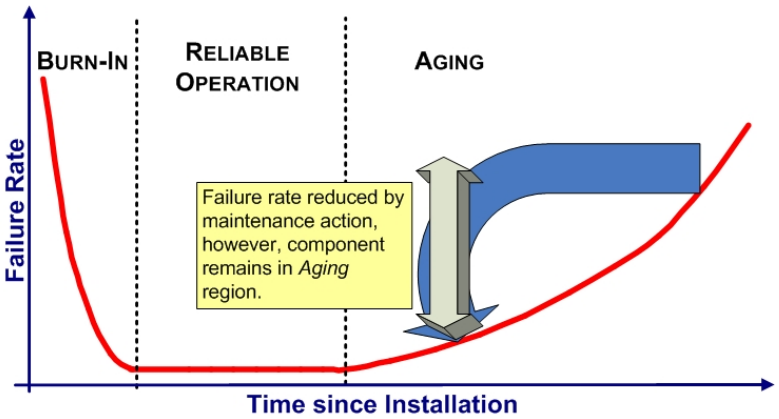


Figure 85: Graphical interpretation of performing maintenance that does not return component(s) to *Reliable Operation* region.

The situation depicted in Figure 85 still produces a benefit for the utility in terms of a reduced failure rate. However, this improvement in reliability is reduced from what could have been achieved had the correct level of maintenance been performed. In programs using more than two action levels, it may also be possible to compute the sensitivity of the failure rate to the cost of maintenance to determine the optimal return on investment. In any case, the net effect of the scenario in Figure 85 is to reduce to the number of avoided failures, thereby, decreasing the effectiveness of the diagnostic program.

9.2.4 Failures Missed During Generation Phase

Diagnostic tests themselves do not generally possess accuracies that are near 100%. This means that a portion of their diagnoses will be incorrect. For a k level diagnostic test the following consequences can result from misdiagnoses:

- If the diagnostic test classifies a component into the k^{th} class when its true condition is less than k, then the component will produce a service failure.
- If the diagnostic test classifies a component into class j and its true condition is class h, where $h>j$, then the component will receive a more expensive corrective action than was needed.

Each of the above consequences leads to a different consequence cost. In the first case, a service failure will occur that will incur the cost of the failure plus any additional customer penalties. In the second case, an unneeded corrective action will be performed that will increase the initial cost of the diagnostic program. One can use set theory to demonstrate these observations. First, it is necessary to define a few additional variables as shown in Table 34.

Table 34: Variable definitions for computing the number of undiagnosed "bad" components in the at-risk population.

Variable	Definition
$\underset{\sim}{X}_B = \bigcup_{i=1}^{k-1} \underset{\sim}{X}_i$	Set of components in the at-risk population that the diagnostic determines will each require a corrective action as they are expected to fail within the target time horizon.
$\underset{\sim}{X}_G = \underset{\sim}{X}_k$	Set of components in the at-risk population that the diagnostic predicts are "good" and so will not fail in the specified time horizon.
$X_B = \left\lvert \underset{\sim}{X}_B \right\rvert$	The number of components diagnosed as "bad" in the at-risk population.
$X_G = \left\lvert \underset{\sim}{X}_G \right\rvert$	The number of components diagnosed as "good" in the at-risk population.
$\underset{\sim}{X}_b$	Set of components in the at-risk population that are "bad" and will fail in the near future.
$\underset{\sim}{X}_g$	Set of components in the at-risk population that are "good" and so will not fail in the near future.
$X_b = \left\lvert \underset{\sim}{X}_b \right\rvert$	The number of "bad" components in the at-risk population.
$X_g = \left\lvert \underset{\sim}{X}_g \right\rvert$	The number of "good" components in the at-risk population.

The number of overlooked "bad" components is computed from set theory according to:

$$F_{UD} = \left\lvert \underset{\sim}{X}_b \cap \underset{\sim}{X}_G \right\rvert, \tag{9.8}$$

where,

F_{UD} = number of "bad" components that were incorrectly diagnosed as "good."

Therefore, the total number of failures (ΔF) that will be avoided during the specified time horizon by the diagnostic program in the at-risk population is:

$$\Delta F = F_D - F_{UD},$$
$$\Delta F = \left|\underset{\sim}{X}_b \cap \underset{\sim}{X}_B\right| - \left|\underset{\sim}{X}_b \cap \underset{\sim}{X}_G\right|, \quad \textbf{(9.9)}$$

where,

ΔF = Net number of avoided failures [Failures],

F_D = Number of diagnosed failures [Failures],

F_{UD} = Number of undiagnosed failures [Failures].

Equation (9.9) requires some explanation. The first term corresponds to the number of "bad" components that were correctly diagnosed by the diagnostic as "bad." The second term, on the other hand, refers to the number of "bad" components that were incorrectly diagnosed as "good." These components would not have received the corrective action that was needed. These incorrect diagnoses reduce the program's potential impact of the reliability of the at-risk population.

Returning to the example presented earlier in Figure 82, Figure 86 shows a possible classification of the at-risk population using a diagnostic test.

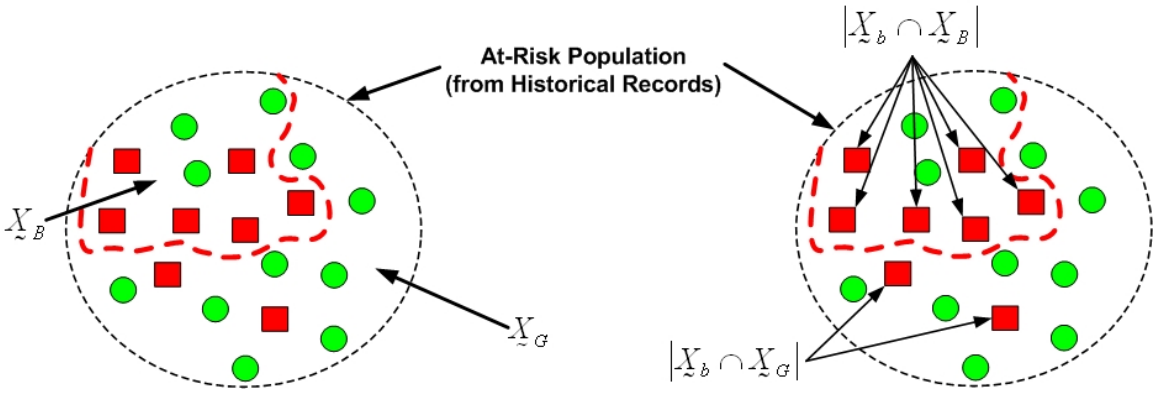


Figure 86: Sample results of diagnostic testing and partitioning of the at-risk components into $\underset{\sim}{X}_B$ and $\underset{\sim}{X}_G$ subsets. Squares (■) represent components that will fail while the dots (●) represent those that will not.

Using Figure 86, ΔF can then be described by the following:

$$\begin{aligned} \Delta F &= \left|\underset{\sim}{X}_b \cap \underset{\sim}{X}_B\right| - \left|\underset{\sim}{X}_b \cap \underset{\sim}{X}_G\right|, \\ &= 6 - 2 = 4 \text{ Failures}, \\ \text{Yield} &= \frac{4}{19} = 0.211 \left[\text{Failures} / \text{Test}\right]. \end{aligned} \tag{9.10}$$

According to the above scenario, the utility would experience a net savings of four failures as a result of its diagnostic program. This translates to a yield of 0.211 [Failures/Test]. Furthermore, the number of corrective actions required to achieve this reduction is:

$$\begin{aligned} X_b = 8 \Rightarrow \frac{X_b}{X_{AR}} &= \frac{8}{19} = 0.421 \left[\text{Maintenance Actions} / \text{Test}\right], \\ \frac{\Delta F}{X_b} &= \frac{4}{8} = 0.500 \text{ [Failures / Maintenance Action]}. \end{aligned} \tag{9.11}$$

Therefore, the scenario in Figure 86 requires that corrective action be performed on 42.1% of the at-risk population. This translates into a reduction in failures of 0.5 [Failures/Corrective Action]. On the other hand, had the utility chosen to act on the entire at-risk population, the following results would have been obtained:

$$\begin{aligned} \Delta F &= \underset{\sim}{X}_b \\ &= 8 - 0 = 8 \text{ Failures}, \\ \frac{\Delta F}{X_{AR}} &= \frac{8}{19} = 0.421 \text{ [Failures / Corrective Action]}. \end{aligned} \tag{9.12}$$

This data shows that a greater number of failures would have been avoided by performing maintenance on the entire population. However, each corrective action would avoid only 0.421 failures as compared to 0.500 failures with the diagnostic program.

One may also compute the accuracies, both overall and condition-specific, for this example. These results are shown in Table 35.

Table 35: Summary of accuracies for the at-risk population of components shown in Figure 86.

Group	Correct Diagnoses [#]	Incorrect Diagnoses [#]	Correct Diagnoses [%]	Incorrect Diagnoses [%]
"Good"	8	2	80.0%	20.0%
"Bad"	6	3	66.7%	33.3%
Overall	14	5	73.7%	26.3%

Note that the overall accuracy of the diagnostic was 73.7%. Even with this low level of accuracy the diagnostic program was able to avoid more failures per corrective action than the complete replacement approach. It must be noted that this information will not be available to the utility for at least two years after testing is completed. However, after this time, the utility will be able to determine the actual number of failures it was able to avoid through comparison with the original predictions.

Each of the "missed" failures contributes an additional cost component to the total cost defined in Section 9.1.5. This addition is shown in Figure 87.

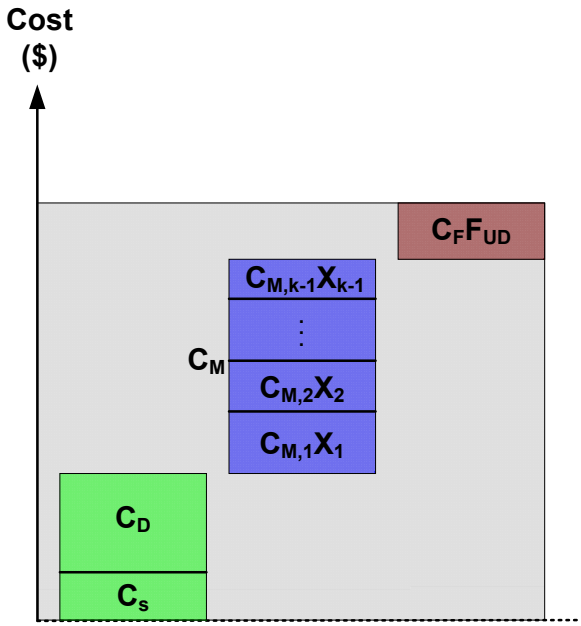


Figure 87: Updated cost diagram illustrating the diagnostic program costs including diagnostics (C_D & C_S), maintenance (C_M), and undiagnosed failures ($C_F \cdot F_{UD}$).

The following sections will explore this concept and its relation to the overall economic savings.

9.3 ECONOMIC BENEFIT

The diagnostic program can produce financial savings for a utility by avoiding future failures in the at-risk population and by performing only the necessary corrective actions. The economic benefit is derived from the resulting cost difference between the diagnostic program and an alternate program. Examples of alternate programs include another diagnostic program, complete replacement of the at-risk population, and "run to failure." This last alternative, "run to failure," represents the standard approach taken by utilities for managing their aging equipment. The "run to failure" approach simply means that components are repaired or replaced only once they have failed in service. This approach will be of great interest in Chapter 10. For now, the objective is to explore the properties of benefit calculations.

Section 9.3.1 presents the total cost function for a general diagnostic program. In sections 9.3.2 and 9.3.3, the savings resulting from this general diagnostic program as compared to complete replacement and "run to failure" are examined.

9.3.1 Total Cost of a General Diagnostic Program

Section 9.1 and Section 9.2 demonstrated the calculation steps needed to compute the total cost of a diagnostic program over a period of T_H years. The same process in the following basic form:

$$C_{Total}^{DP} = X_{AR}\left(C_T + C_S\right) + \sum_{i=1}^{k} C_{M,i} X_i + F_R X_{AR} T_H \left(1-P\right)\left(C_{M,1} + N \cdot C_{Cust}\right), \tag{9.13}$$

where,

C_{Total}^{DP} = Total cost of diagnostic program [$].

F_R = Average failure rate of at-risk population [Failures/Component/Year],

T_H = Target time horizon [Years],

P = Overall accuracy of diagnostic test.

Note that (9.13) can be decomposed into three components:

$$C_{Total}^{DP} = \underbrace{X_{AR}\left(C_T + C_S\right)}_{\text{Diagnostic Testing}} + \underbrace{\sum_{i=1}^{k} C_{M,i} X_i}_{\text{Corrective Actions}} + \underbrace{F_R X_{AR} T_H \left(1-P\right)\left(C_{M,1} + N \cdot C_{Cust}\right)}_{\text{Consequence}}. \tag{9.14}$$

Sections 9.3.2 and 9.3.3 demonstrate the use of (9.13) in calculating economic benefits.

9.3.2 Complete Replacement Program

This section demonstrates the economic savings a utility could obtain from a diagnostic program as compared to a complete replacement program. The total cost of a complete replacement program is:

$$C_{Total}^{CR} = C_{M,1} \cdot X_{AR}, \tag{9.14}$$

where,

$C_{M,1}$ = Total cost to replace a component [$/Component],

X_{AR} = Number of components in the at-risk population [Components].

The savings that a diagnostic program would produce is computed as:

$$S = C_{Total}^{CR} - C_{Total}^{DP}$$

$$= C_{M,1}X_{AR} - X_{AR}\left(C_T + C_S\right) - \sum_{i=1}^{k} C_{M,i}X_i - F_R X_{AR} T_H \left(1-P\right)\left(C_{M,1} + N \cdot C_{Cust}\right) \quad \textbf{(9.15)}$$

Rearranging the terms slightly in (9.15) leads to:

$$S = C_{M,1}\left(X_{AR} - X_1\right) - X_{AR}\left(C_T + C_S\right) - \sum_{i=2}^{k} C_{M,i}X_i - F_R X_{AR} T_H \left(1-P\right)\left(C_{M,1} + N \cdot C_{Cust}\right) \quad \textbf{(9.16)}$$

As in (9.14), two components to the savings can be readily seen:

$$S = \underbrace{C_{M,1}\left(X_{AR} - X_1\right)}_{\text{Corrective Action Savings}} \underbrace{- X_{AR}\left(C_T + C_S\right) - \sum_{i=2}^{k} C_{M,i}X_i - F_R X_{AR} T_H \left(1-P\right)\left(C_{M,1} + N \cdot C_{Cust}\right)}_{\text{Diagnostic Program Cost}} \quad \textbf{(9.17)}$$

The first component, corrective action savings, represents the reduction in replacement spending by utilizing the diagnostic program. The remaining terms constitute the remaining cost of the diagnostic program. For there to be a savings, the diagnostic program cost must be less than the corrective action savings. This implies that as compared to the complete replacement scenario, the diagnostic program generates its savings from reduced spending on corrective actions. Figure 88 shows the savings in the cost diagram format.

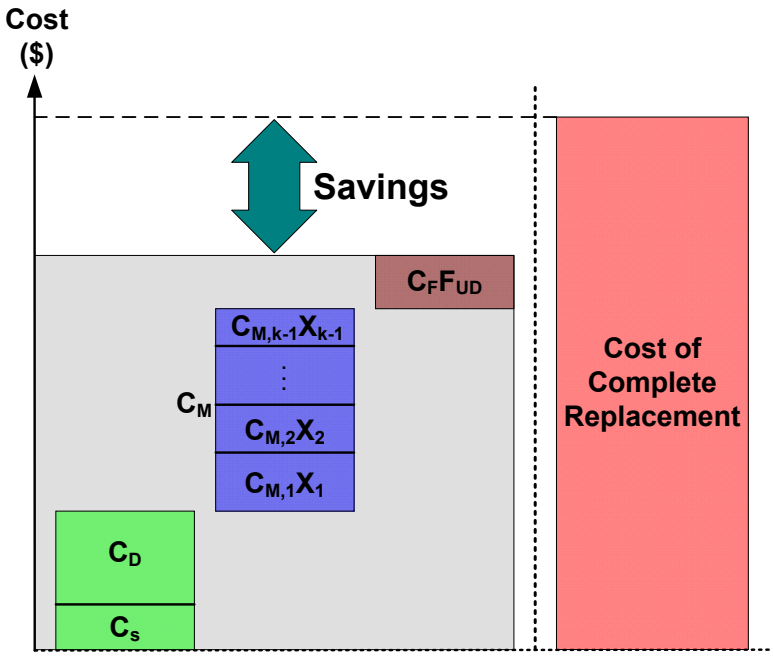


Figure 88: Updated cost diagram illustrating diagnostic program savings as compared to complete replacement of the at-risk population.

Section 9.3.3 demonstrates a second alternative program that will show a different source of savings.

9.3.3 "Run to Failure" Program

The "run to failure" program can be approached in a similar way as the complete replacement case in Section 9.3.2. The total cost of a "run to failure" program can be defined as:

$$C_{Total}^{RF} = F_R X_{AR} T_H \left(C_{M,1} + N \cdot C_{Cust} \right) \tag{9.18}$$

where,

C_{Total}^{RF} = Total cost of the "run to failure" program [$].

Similar to the complete replacement case, the cost difference between the diagnostic program and the "run to failure" program is computed as:

$$S = C_{Total}^{FR} - C_{Total}^{DP},$$

$$= F_R X_{AR} T_H \left(C_{M,1} + N \cdot C_{Cust}\right) - X_{AR}\left(C_T + C_S\right) - \sum_{i=1}^{k} C_{M,i} X_i - F_R X_{AR} T_H \left(1-P\right)\left(C_{M,1} + N \cdot C_{Cust}\right). \quad \textbf{(9.19)}$$

Rearranging the terms slightly in (9.19) leads to:

$$S = F_R X_{AR} T_H P\left(C_{M,1} + N \cdot C_{Cust}\right) - X_{AR}\left(C_T + C_S\right) - \sum_{i=2}^{k} C_{M,i} X_i. \quad \textbf{(9.20)}$$

As in (9.14), two components to the savings can be readily seen:

$$S = \underbrace{F_R X_{AR} T_H P\left(C_{M,1} + N \cdot C_{Cust}\right)}_{\text{Reliability Savings}} - \underbrace{X_{AR}\left(C_T + C_S\right) - \sum_{i=2}^{k} C_{M,i} X_i}_{\text{Diagnostic Program Cost}}. \quad \textbf{(9.21)}$$

Note that in this example that the savings component of (9.21) is now the result of improved reliability rather than reduced spending on corrective actions. Once again the diagnostic program produces a savings when the diagnostic program cost is less than the reliability savings it produces.

The two programs used in the examples demonstrate the two extreme cases that can occur: (1) savings come exclusively from reduced spending on corrective actions as in the complete replacement example and (2) savings come exclusively from improved reliability as in the "run to failure" example. Comparing two diagnostic programs would likely produce a mix between reliability and correct action savings.

Unfortunately, there is an additional complexity that must be considered in an economic benefit analysis: uncertainty. Section 9.4 introduces the concept of "value at risk" for addressing this issue.

9.4 BENEFIT CALCULATIONS

The cost facet presented thus far can be easily used to compute the total cost of a diagnostic program if all the input variables are deterministic values. If the value of each input is known then the calculation of the total cost is straightforward. Unfortunately, this

approach cannot be used for a "real" diagnostic program as the assumed deterministic nature of the inputs is incorrect. The inputs and their uncertainty must be, therefore, treated as random variables.

In the examples presented in Section 9.3, the cost difference between any two programs could be computed directly. Now, the results of this operation will take the form of a distribution in which there are cases where the diagnostic program will be more expensive and thus produce a loss. In other cases, the diagnostic program will cost less and so produce a savings. By analyzing this distribution, one can extract two key measures of the likelihood that the diagnostic program will be successful:

- Value at risk – The maximum loss that would be incurred by 95% (or any desired level of confidence) of the cases.
- Probability of loss – The expected probability that a diagnostic program would generate a loss or less than zero savings.

These measures are each discussed in detail in the following sections.

9.4.1 Value at Risk

The term value at risk refers specifically to a kind of worst case performance of the diagnostic program as compared to any other program. A typical confidence level is 95%. The resulting value at risk is the dollar figure that corresponds to the 95% confidence level as illustrated in Figure 89.

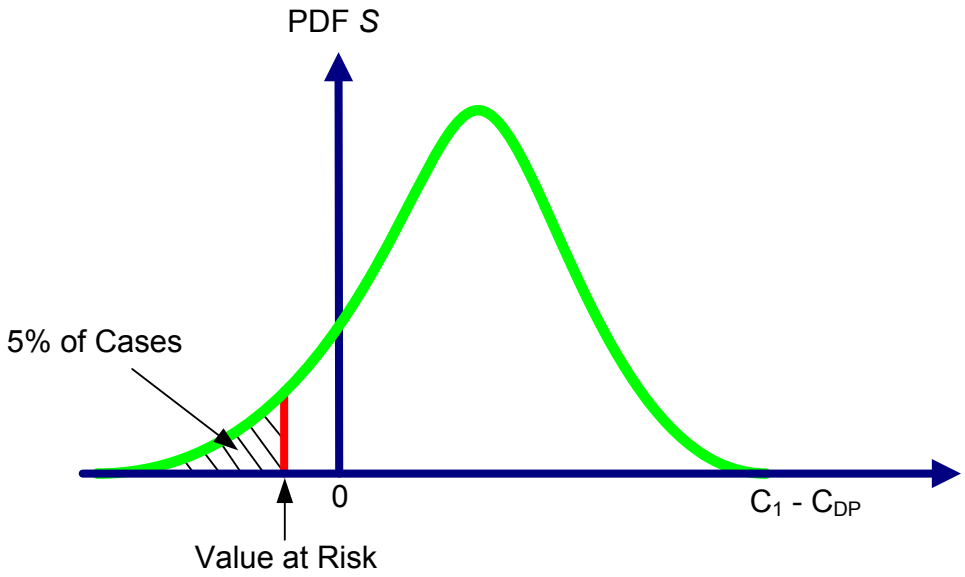


Figure 89: Illustration of the definition of value at risk.

As Figure 89 shows, the value at risk provides a lower bound for 95% of the cases. An alternative approach is to utilize the probabilities of savings and loss as discussed in Section 9.4.2.

9.4.2 Probabilities of Savings and Loss

The probabilities of savings and loss are simple measures of how much of the savings distribution is located on either side of the y-axis. Figure 90 and Figure 91 show the definitions of probability of savings and loss, respectively.

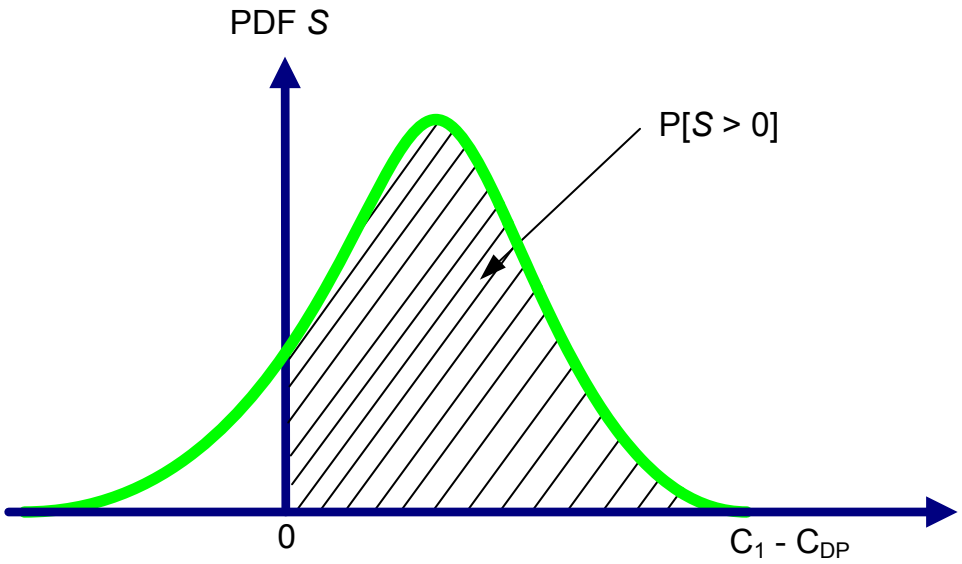


Figure 90: Graphical definition of the probability of savings.

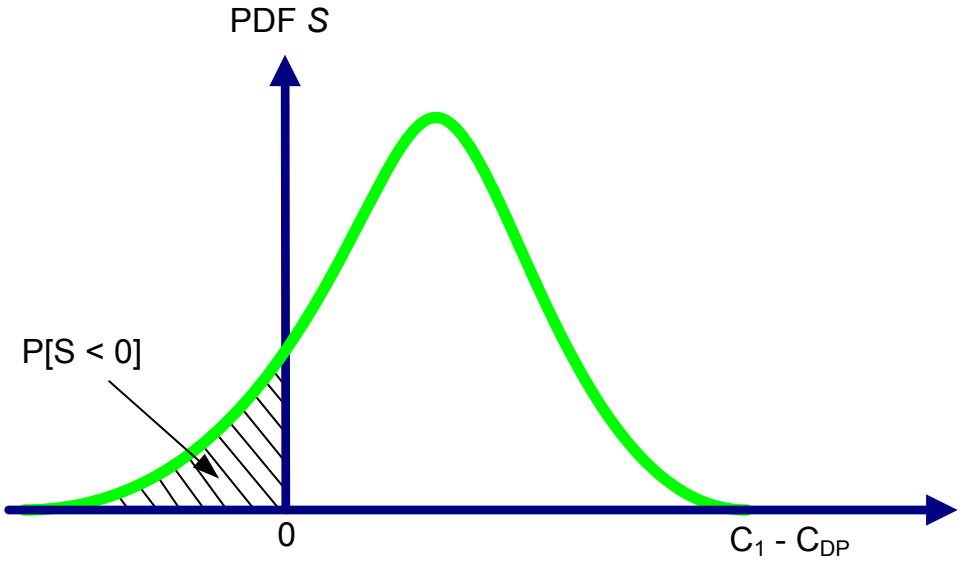


Figure 91: Graphical definition of the probability of loss.

Using value at risk and probability of loss (or savings) one can provide a measure of how likely the diagnostic program is to generate a loss. At the end of the day, the goal is to maximize the probability that the diagnostic program will generate a savings.

9.4.3 Stochastic Optimization

It requires that the cost function described in this chapter be formulated as a stochastic optimization problem in which those parameters with uncertainty are appropriately modeled by distributions as opposed to deterministic values. Using this objective function, it would then be possible to maximize the savings or to simply minimize the costs of the diagnostic program. It may also be of interest to maximize the expected system reliability or the efficiency of each maintenance action. There are numerous objectives that a utility may wish to pursue. However, the goal here is to cast the problem in such a way that it may be solved using the available optimization techniques. Therefore, identification of these techniques and their associated assumptions and requirements is vital to the formulation.

One may approach this problem based on two possible cases. The first is a linear problem in which the total savings, S_T, can be written as the summation of k random variables as:

$$S_T = S_0 + S_1 + S_2 + ... + S_k, \tag{9.22}$$

where,

S_T = Total savings,

S_0 = Summation of all deterministic quantities,

$S_i,\ i \in \{1, 2, ..., k\}$ = Random variables representing different savings components with known PDFs, f_i(x).

As a result of this formulation, the distribution for S_T would be the k-fold convolution of the PDF of each individual component. This process may be done recursively two variables at a time and noting that $Y_1 = S_1 + S_2$. The PDF of Y_1 can then be obtained using:

$$f_{Y_1}(y) = \int_0^y f_1(x) f_2(y - x)\, dx \tag{9.23}$$

This may then be repeated considering two variables at a time by noting that $Y_2 = Y_1 + S_3$ and so on to obtain the PDF of the combined sum. As mentioned, this process may be performed on a linear system of random variables. Unfortunately, things become more complicated when the summation includes terms that are nonlinear combinations of random variables such as:

$$S_T = S_0 + S_1 S_2 + \ldots + S_k \tag{9.24}$$

As (9.13) shows, the objective function for the savings of a diagnostic program is of the form shown in (9.24).

One technique for handling nonlinear functions of random variables is to linearize the function around the expected values of each of the random variables. This process would allow one to obtain a solution. However, this solution would only be valid in the vicinity of the point(s) around which the linearization has been performed. Other techniques such as a Markov decision model may be employed as well [40], [42], [44].

Another, less rigorous technique, is to employ a Monte Carlo simulation in order to construct the distribution of the savings directly for specific programs under comparative analysis. This technique shall be adopted in the simulation studies presented in Chapter 10.

9.5 SUMMARY

This chapter has discussed the cost functions needed to compute the total cost of a general diagnostic program. In addition, several key issues were discussed that reduce the effectiveness of the diagnostic program since they result in reduced economic benefit. To determine the economic benefit of a diagnostic program, the total program cost must be compared to an alternative program. Several alternative programs may be used but they typically fall under one of the following classes: (1) alternate diagnostic program, (2) complete replacement, and (3) “run to failure”. In the cases of complete replacement and

"run to failure," the savings functions have been derived. These show that these two cases represent two extremes in that the savings as compared to a complete replacement program is the result of reduced spending on corrective actions while the savings as compared to a "run to failure" program are the result of improved reliability. In the case of an alternate diagnostic program, the savings would likely be a combination of reduced corrective action spending and improvements in reliability.

This chapter has also discussed the issue of uncertainty inherent in the cost functions themselves. To properly assess the savings function it is necessary to perform a stochastic simulation that would then allow one to examine the value at risk and probabilities of loss or savings. These techniques are extensively employed in Chapter 10 for the analysis and simulation of several economic case studies.

CHAPTER 10: SIMULATION STUDIES

The focus of this chapter is to demonstrate the effect different case studies have on the likelihood of obtaining economic savings. To that end, this chapter will illustrate the model described in Chapter 9 using several stochastic simulations. Several case studies will be illustrated and are based on the selection of different sizes of at-risk populations and on different region types. In the case of the former, the at-risk population size will be chosen to be either 100 or 1000. Each case will be demonstrated using the region definitions shown in Table 36. These are necessary in order to show the span of results that could be encountered within a typical utility's distribution system. Transmission class equipment would involve larger numbers of customers and, in fact, would involve customers outside the utility's operating region. Note that these are based solely on the average number of customers that would be impacted by the failure of one component.

Table 36: The average number of customers that would be impacted by a failure of one component for each region type.

Region Type	Average Customers [#]
Rural	20
Suburban	200
Urban	20,000

The stochastic simulation setup is described in Section 10.1. Since the interest is in demonstrating the effects on savings, it is necessary to define the base case to which a comparison can be made. This base case is described in Section 10.2. Based on the

different at-risk population sizes and region types, a total of six cases will be investigated. The results of these simulations are presented in Sections 10.3 and 10.4.

10.1 SIMULATION SETUP

Table 37 shows the input parameters that are used in each of the simulation studies.

Table 37: Description and ranges for cost parameters needed to estimate the total cost of a diagnostic program.

Cost Component	Input Parameter	Description	Range
Selection	Time Horizon	The time period for which the diagnostic is assumed to be valid.	5 Years
	X_{AR}	Size of at-risk population	100 or 1000 Components
	Failure Rate	Local failure rate of at-risk population.	0.001 – 0.10 [Failures/Comp./Year] 0.1-10 [Failures/100 Comps./Year]
Diagnostic	Diagnostic Test	Total cost of performing diagnostic testing on each segment (includes switching crew if needed).	0.5 Cost Units
	Failure on Test Rate (FOT)	Percentage of segments that fail during diagnostic testing.	2.5%
	Overall Diagnostic Accuracy	Percentage of correct diagnoses made during the time horizon.	51 – 99%
Corrective Action	Installation Cost	Total cost to install a replacement component.	2 Cost Units
Consequence	Average Customers	The average number of customers affected by the failure of one component.	20, 200, or 20000 Customers
	Time of Failure	Day of week and time of day when failure occurs. Outside of normal business hours produces overtime factor.	0 – 168 hours
	Failure Penalty Cost	Total amount utility is charged as a result of service interruptions.	0.1 – 0.5 [Cost Units/Customer/Failure]
	Normal Repair Cost	Cost of crew and parts to repair a segment (does not include impact to customers or reliability indices).	2 Cost Units 2.5 Cost Units (Overtime)

Several of the input parameters listed in Table 37 are treated as random variables while others will be swept through all possible combinations. The selection of inputs is based on uniform distributions with the limits specified in Table 37. Table 38 lists these variables are their corresponding treatment.

Table 38: Description of how each "random" input is treated during simulation.

Input Variable	Treatment
X_{AR}	All Combinations
Failure Rate	All Combinations
Overall Diagnostic Accuracy	All Combinations
Average Customers	All Combinations
Time of Failure	Random
Customer Penalty Rate	All Combinations

Note that during the simulation process itself, the actual number of customers affected by an outage is treated as a random variable that is distributed as a truncated exponential distribution. However, this random variable depends on the average number of customers.

Section 10.2 describes the base case that each of the diagnostic programs will be compared against.

10.2 BASE CASE – "RUN TO FAILURE"

The base case (alternate program) to which all savings calculations will be computed is known as "run to failure". In terms of maintenance, this program represents the corrective maintenance approach. During this program, the number of components deemed to be "bad" (and would ideally be identifiable by a diagnostic test) will each produce a service failure that will impact customers. The number of potential customers depends on the region type (rural, suburban, or urban).

The estimate of the true nature of the component population is based on the local failure rate of the at-risk population at the start of the testing program. Figure 92 shows the percentage of the at-risk population that would be expected to fail during the time horizon for each possible failure rate.

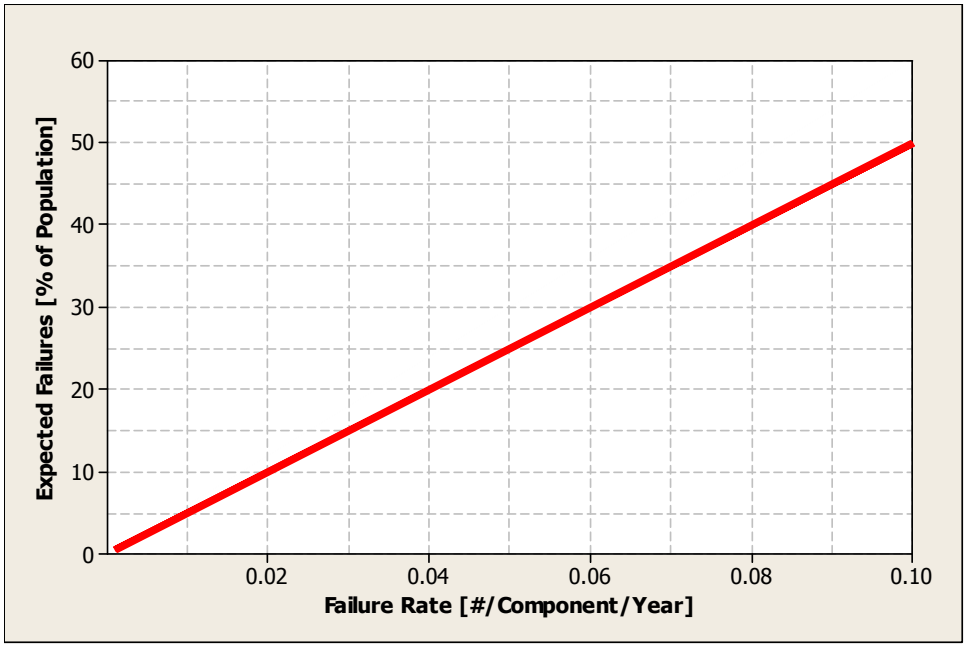


Figure 92: Percentage of "good" components in a population as a function of failure rate [failures/component/year] for a five year time horizon.

From Figure 92, it is straightforward to extract the G/B ratios for a select group of failure rates as shown in Table 39.

Table 39: Component ratios resulting for different failure rates.

Failure Rate [#/Component/Year]	Good/Bad Component Ratio [G/B]
0.01	95/5
0.02	90/10
0.03	85/15
0.04	80/20
0.05	75/25
0.06	70/30
0.07	65/35
0.08	60/40
0.09	55/45
0.10	50/50

Note that in underground cable system diagnostic programs (as seen in the NEETRAC CDFI project), the cable systems under investigation are generally never below 50/50. It is more common to find diagnostic tests being employed in systems that are closer to a G/B ratio of 90/10.

10.2.1 "Run to Failure" Total Cost

The primary means of illustrating results of the stochastic simulation will be a combination of contour and surface plots. Figure 93 shows one contour plot for each of the six cases defined in Section 10.1. Note that a contour plot uses different colors to show regions whose values lie within the same range. For each of the cases in Figure 93, the intervals represented by each color are unique. The key at this point is to observe the similarities and differences in the shapes of the contours. Note that red portions correspond to high total costs while green portion represent lower total costs for the "run to failure" program.

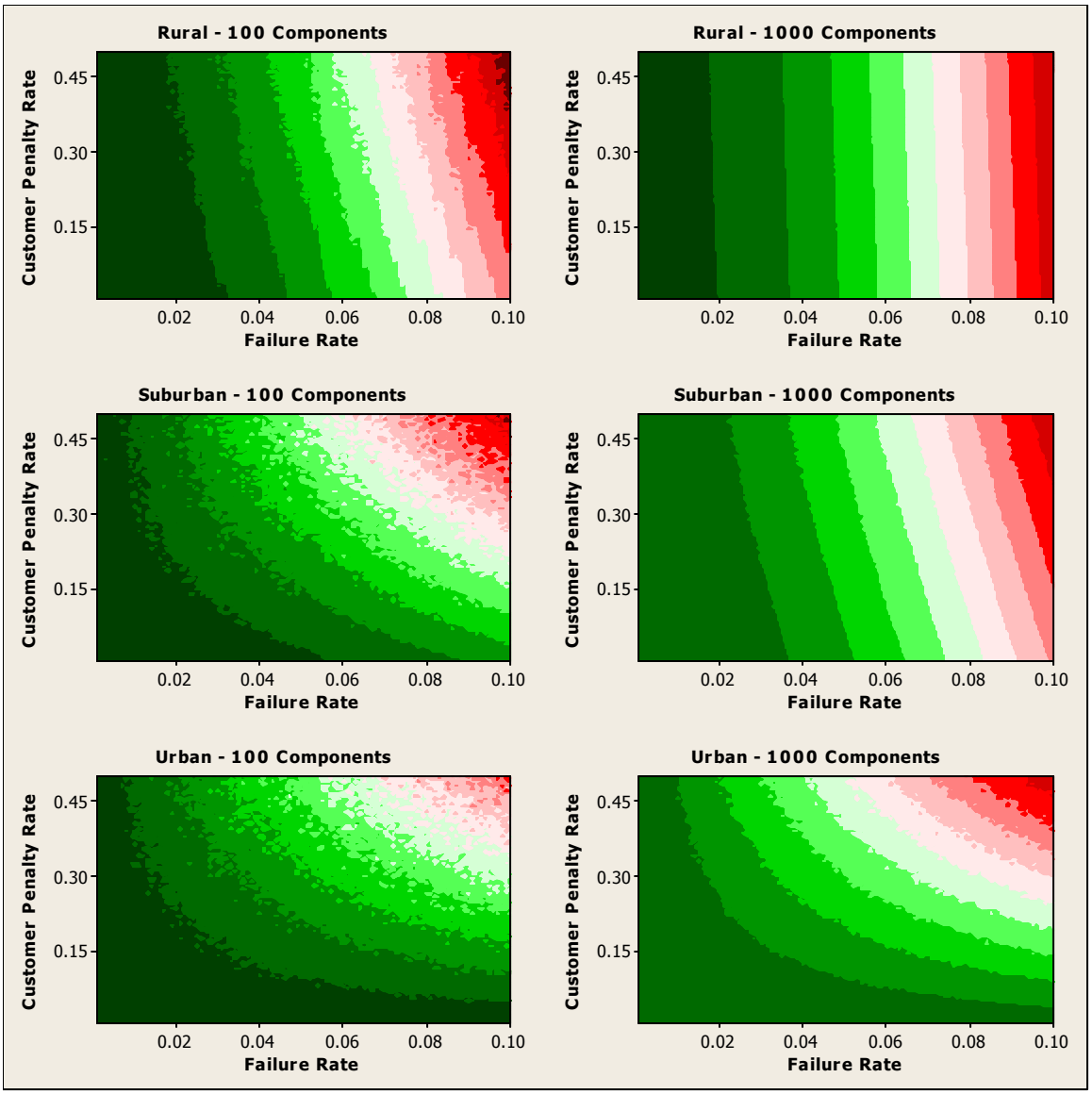


Figure 93: "Run to failure" program cost contours as functions of customer penalty rate [cost units/customer/failure] and failure rate [failures/component/year]. Note that contour intervals are different for each plot. Each region type was used: rural (top row), suburban (middle row), and urban (bottom row). At-risk population sizes of 100 components (left) and 1000 components (right) are also used. Note that any jaggedness at the interface between two contour bands is an artifact of the sampling performed at the start of the simulation.

As Figure 93 demonstrates, there are clear differences in the pattern for each of the scenarios. The following sections explore both the similarities and differences.

10.2.1.1 Rural Cases

If one examines the rural region cases (top row of Figure 93) a clear pattern to the contours can be observed. The contours are parallel to one another in each of the different at-risk population sizes. It is also apparent that as the number of components increases the slope of the contour boundaries increase in magnitude. This implies that the customer penalty is not as influential and, thus, the total cost becomes purely a consequence of the increasing failure rate. Not surprisingly, as the failure rate increases so does the cost of the "run to failure" program. The increasing program cost can also be visualized using a surface plot such as that shown in Figure 94.

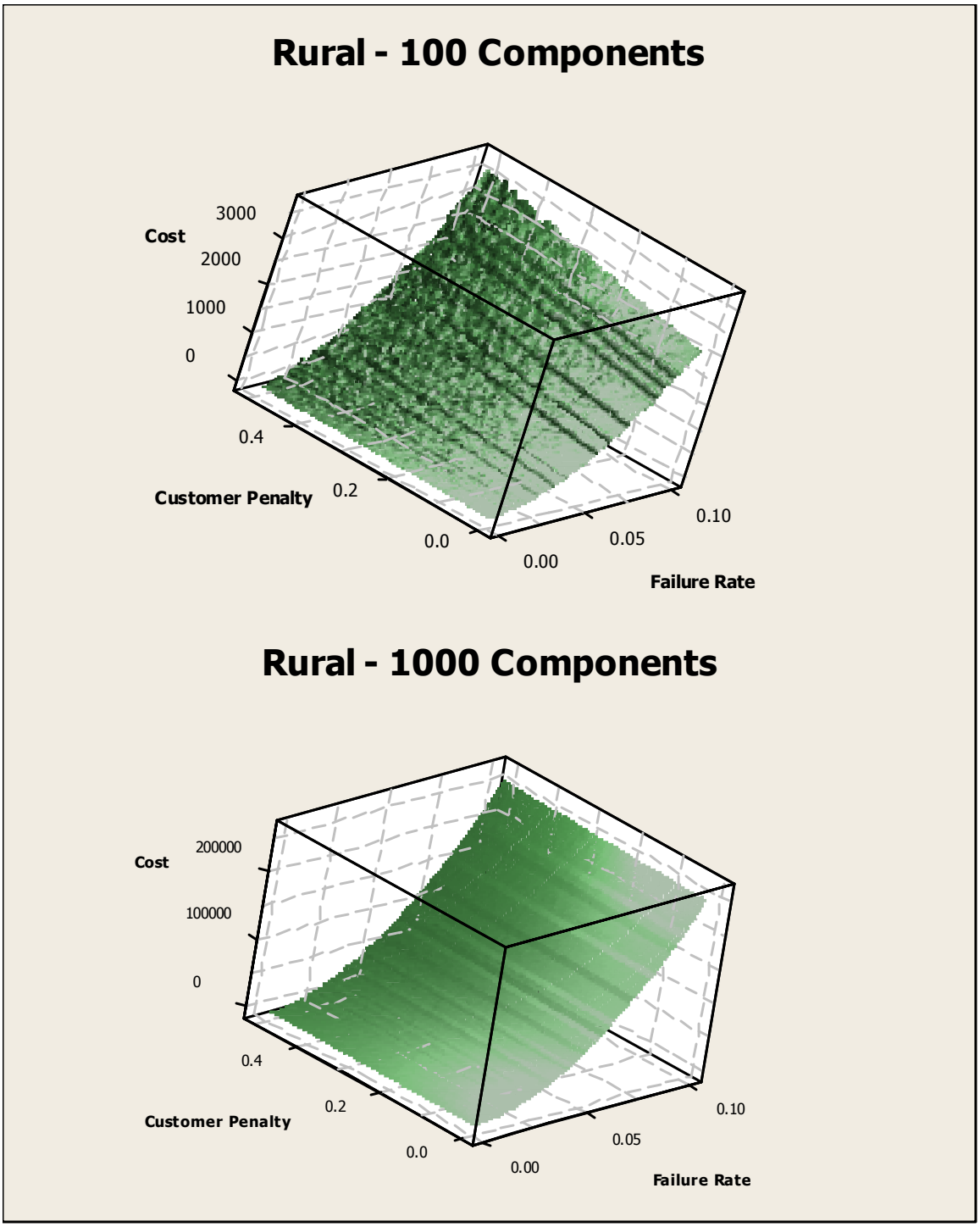


Figure 94: Surface plots of the total cost of "run to failure" as a function of customer penalty rate [cost units/customer/failure] and failure rate [failures/component/year] for a rural region. Note that the jagged surface features are an artifact of the sampling strategy employed and the small number of at-risk components.

Note that for each component population size, the surface is virtually identical but scaled to the corresponding population size.

Figure 94 illustrates that the transition from the low to high cost region is, loosely speaking, smooth and quadratic in shape.

10.2.1.2 Suburban Case

It is also interesting to examine the behavior of the suburban region examples. The resulting surface plots are shown in Figure 95 for both sizes of at-risk population.

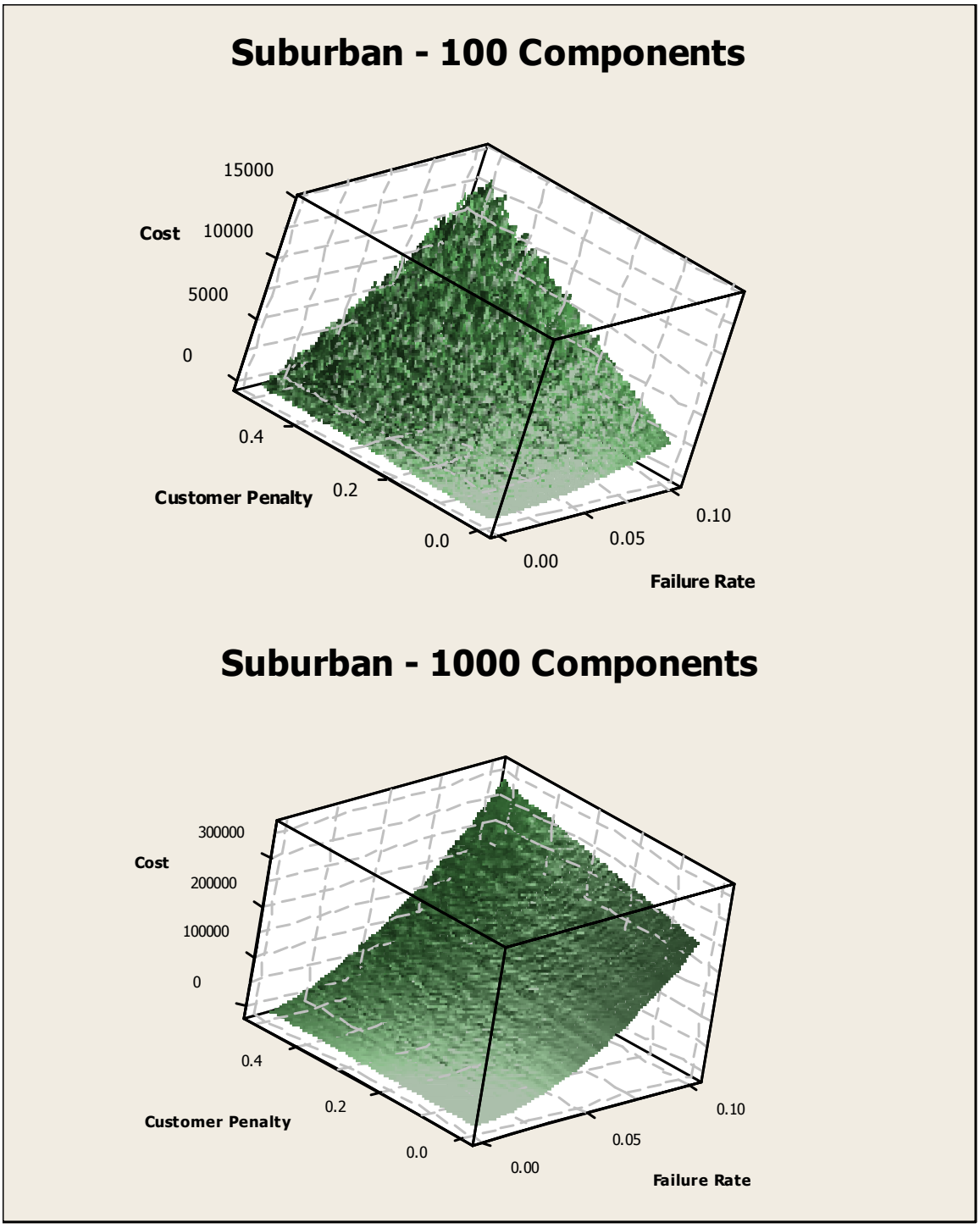


Figure 95: Surface plots of the total cost of "run to failure" as a function of customer penalty rate [cost units/customer/failure] and failure rate [failures/component/year] for a suburban region. Note that the jagged surface features are an artifact of the sampling strategy employed.

The picture is much the same for a population size of 1000 between the suburban and rural region types as shown in Figure 94 and Figure 95. However, as also illustrated by Figure 93, the suburban region with a population of 100 components behaves substantially different from the other cases shown thus far. In this case, the customer penalty rate appears to be much more influential since at the highest failure rate a clear maximum is seen at the maximum customer penalty rate. This is more characteristic of the urban region scenarios as hinted at in Figure 93 and may be a result of the relatively small component population combined with a larger number of customers.

10.2.1.3 Urban Region

The same surface plot examination for the urban region is shown in Figure 96.

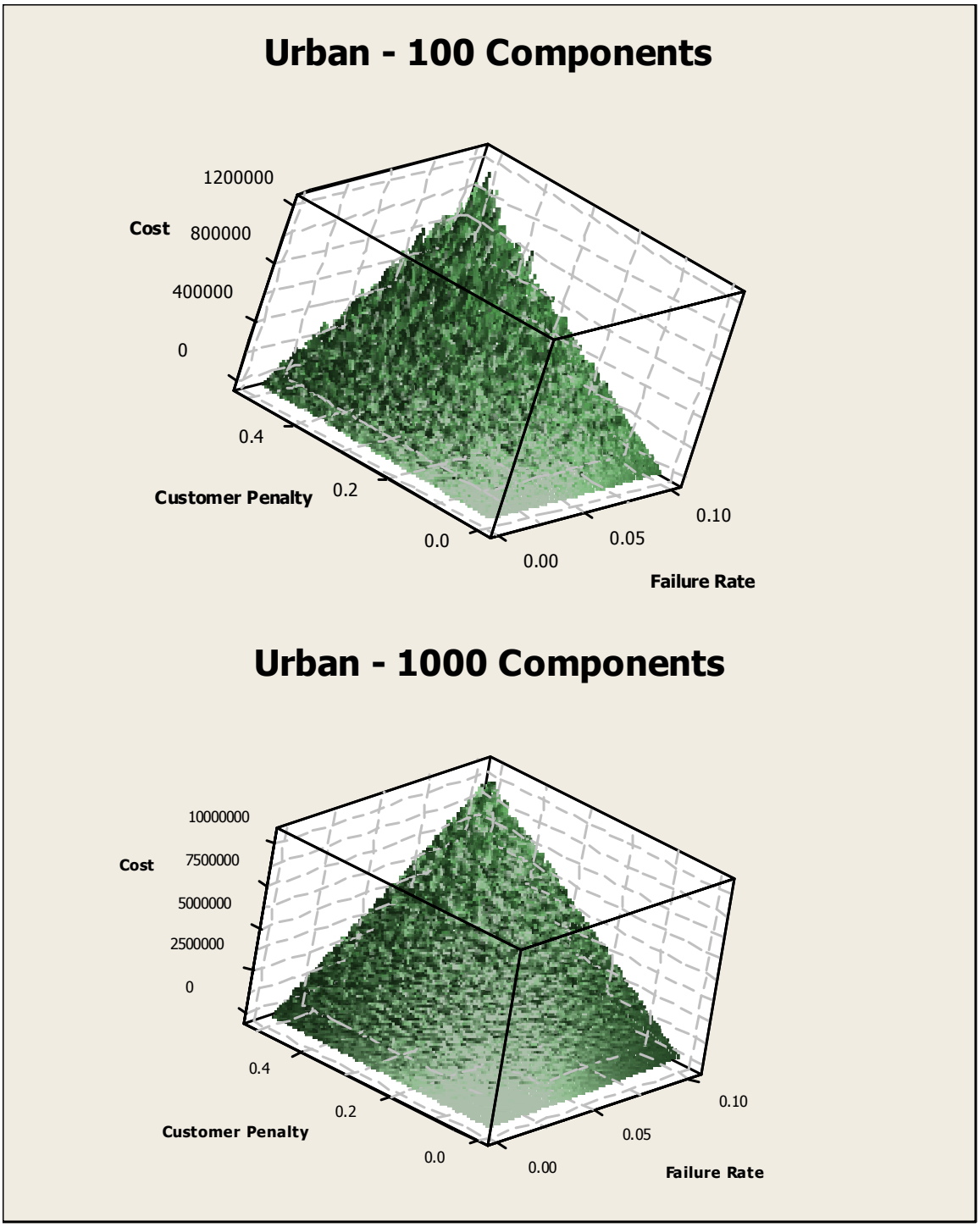


Figure 96: Surface plots of the total cost of "run to failure" as a function of customer penalty rate [cost units/customer/failure] and failure rate [failures/component/year] for an urban region. Note that the jagged surface features are an artifact of the sampling strategy employed.

The surface plots in Figure 96 show the same characteristic as the suburban case with 100 components did in Section 10.2.1.2. Clearly, for the customer penalty rate to be a significant factor, both the penalty itself and the number of customers must be large enough. With the urban region, both criteria are easily met.

10.2.2 Population Sizes

The contour plots shown in Figure 93 demonstrate that there is also an effect that is the result of changes in the at-risk population size. By increasing the population the effect appears to be a reduction in the influence of the customer penalty rate. The most extreme example of this effect is seen in the suburban region examples. In this case, increasing the population size tends to return the contour lines to a parallel configuration as was seen with both rural examples. With the reduced population size, the suburban example contour lines appear nonlinear but still evenly spaced. This is identical to the examples for an urban region. Therefore, for the simulation examples studied in this chapter, there appears to be a crossover point in which the suburban examples appear to be like the other region types depending on the size of the at-risk population.

The following section explores the effect of an at-risk population size on the potential savings.

10.3 SMALL AT-RISK POPULATION (100 COMPONENTS)

This section investigates the economic modeling of a diagnostic program that includes 100 components. The previous section demonstrated the behavior of the total cost for the "run to failure" program. This section will demonstrate the effect the region types have on the potential savings that a 100 component population can deliver. Section 10.3.1 addresses the effects of the failure rate and diagnostic accuracy on the probability of savings. Section 10.3.2 and 10.3.3 similarly examine the effects of customer penalty rate both with failure rate and diagnostic accuracy.

10.3.1 Failure Rate and Diagnostic Accuracy

The failure rate of the at-risk population and the diagnostic accuracy are both factors that the asset manager can select (or specify) at the start of the program. The average failure rate is set by the choice of at-risk population while the diagnostic accuracy is the result of the chosen diagnostic technique. These are vital components to the total cost of the diagnostic program as shown in Chapters 9 and 10.

10.3.1.1 Rural Region

Figure 97 shows a contour plot of the probability of achieving savings as a function of the failure rate and diagnostic accuracy for a rural region.

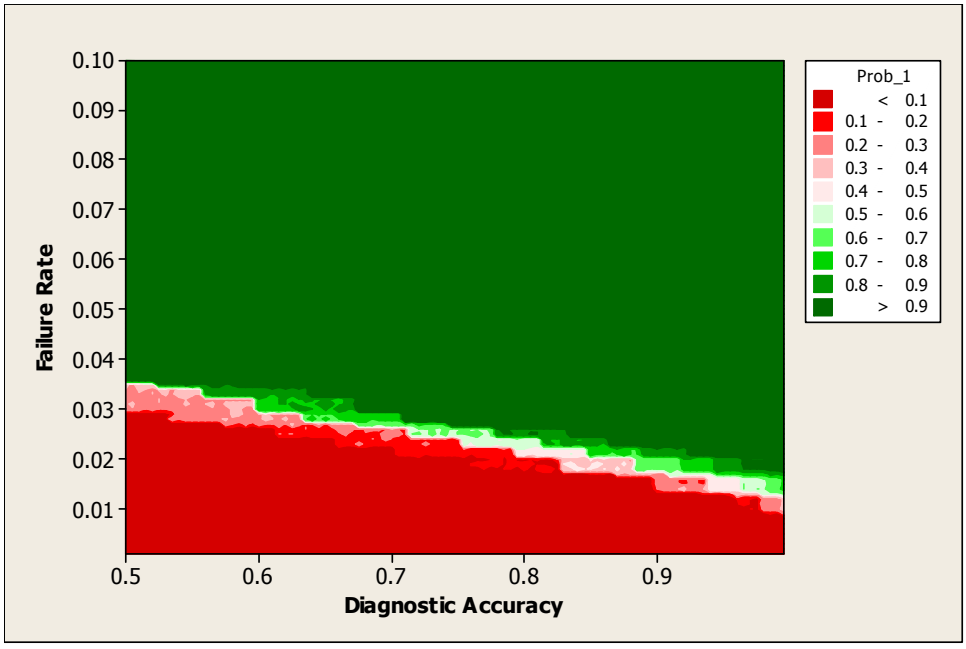


Figure 97: Contour plot of the probability of obtaining savings as a function of the failure rate [failures/component/year] and diagnostic accuracy for a rural region.

Figure 97 clearly shows regions where the probability of obtaining a cost savings as compared to the "run to failure" program is less than 10%. Furthermore, there are failure

rates for which even a 100% accurate diagnostic technique will not yield a savings. Equally, once the failure rate exceeds approximately 0.04 [failures/component/year] (or 4 [failures/100 components/year]) any diagnostic with an accuracy of more than 51% can produce a savings with at least 90% probability. Within the five year horizon of this simulation, the total number of expected failures at this failure rate is approximately 20 failures. This translates to a G/B ratio of 80/20. As long as the population is in worse condition than this, the diagnostic program will produce a savings with probability greater than 90%. Figure 98 shows two of the distributions represented in the contour plot of Figure 97. Note that one case consists entirely of data points that do not generate a savings while the second is composed almost entirely of data points which do generate savings.

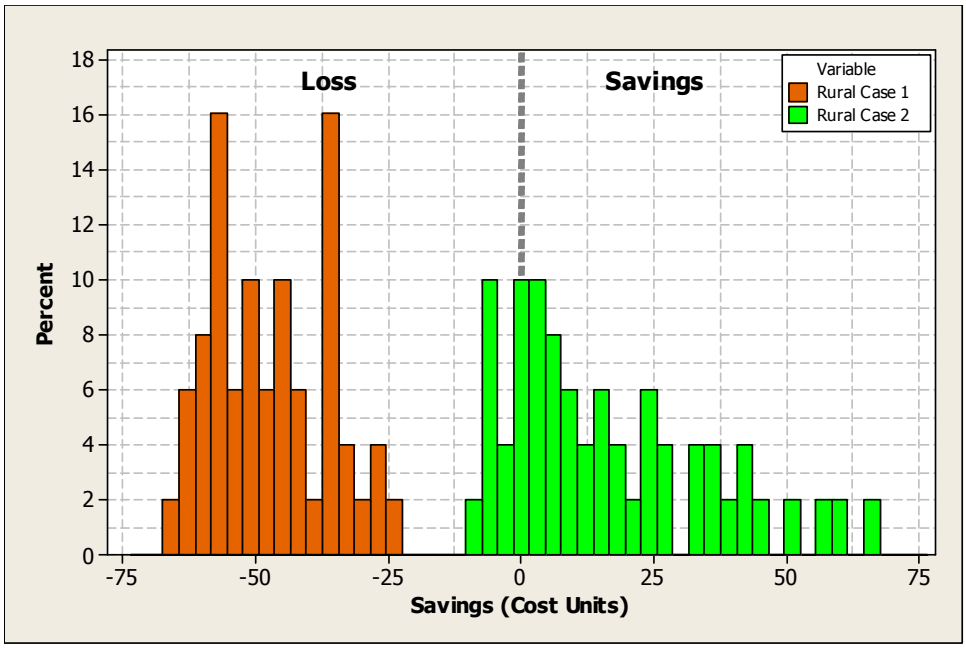


Figure 98: Example distributions for given failure rates and diagnostic accuracies. Case 1 corresponds to a failure rate of 0.02 [failures/component/year] and diagnostic accuracy of 0.8. Case 2 corresponds to a failure rate of 0.025 [failures/component/year] and diagnostic accuracy of 0.8.

By utilizing a more accurate diagnostic, the minimum failure rate needed to guarantee a 90% probability of a net program savings is less than that required for the lower accuracies. Figure 99 shows a close up view of the transition between the low probability contours (less than 0.3) to the high probabilities of savings (greater than 0.7).

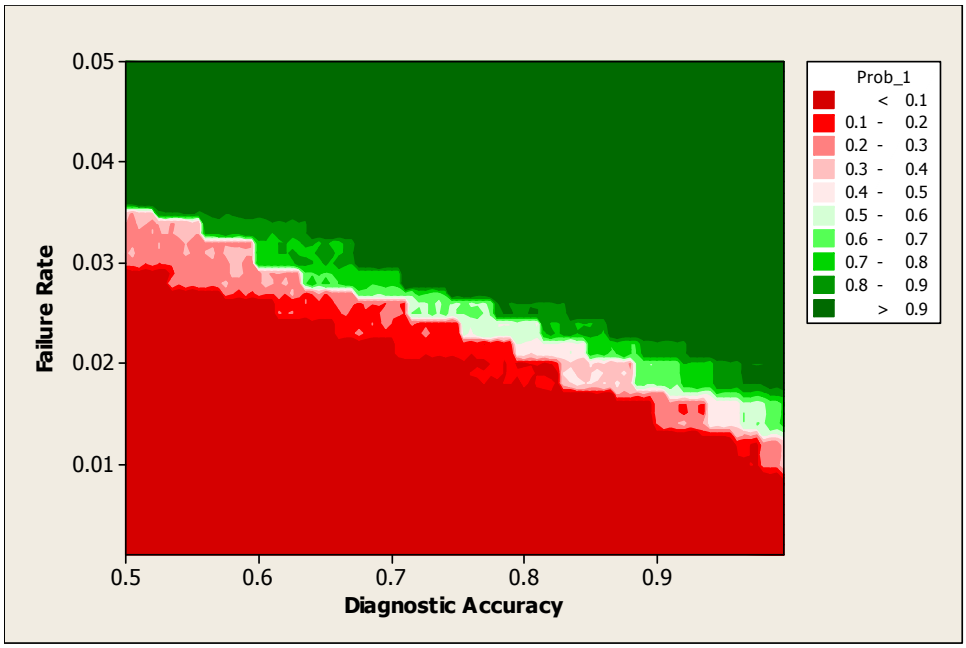


Figure 99: Detailed view of Figure 97 showing the probability contours as a function of failure rate [failures/component/year] and diagnostic accuracy. Note that the jagged surface features are an artifact of the sampling strategy employed.

According to Figure 99, the same probability of savings can be obtained with a failure rate of approximately 0.02 [failures/component/year] (2 [failures/100 components/year]) and a diagnostic accuracy greater than 95%. As a result, one could utilize an increased G/B ratio of 90/10 with the more accurate diagnostic techniques while still achieving the same probability of savings.

From the numerical perspective, it is also useful to examine the relative sizes of the probability ranges shown in Figure 97 and Figure 99. Table 40 shows the resulting percentages.

Table 40: Percentage of data shown in Figure 97 within each of the probability ranges.

Probability Range	Count [% of Total Cases]
< 0.1	19.83
0.1 – 0.2	1.37
0.2 – 0.3	1.56
0.3 – 0.4	0.72
0.4 – 0.5	0.48
0.5 – 0.6	0.61
0.6 – 0.7	0.61
0.7 – 0.8	0.97
0.8 – 0.9	1.10
> 0.9	72.75

Note that the percentages shown in Table 40are based on the specified probability range. The percentages would likely change if the probability ranges are altered.

As Table 40 shows, the vast majority of cases shown in Figure 97 (73%) have a probability of savings that is greater than 90%. This is a result of the G/B ratio. On the other hand, approximately 20% of the cases yield probabilities of savings that are less than 10%. Again, this is the result of a low failure rate. In between the two extremes lies the remaining few percent of the cases. It seems quite clear from this analysis that the probability of savings is largely dependent on the failure rate. In cases where the failure rate is not high enough, the diagnostic accuracy can make some difference.

It is worth noting that the average probability of loss can be determined using the data shown in Table 40. The calculation process is illustrated in Table 41. In this table, the "worst case" and "best case" approaches are based on the low and high probabilities, respectively, that define ach contour's range. The weighting refers to the contribution each contour makes based on the fraction of all cases within the contour to the average probability of loss. In other words, the average probability of loss is a weighted average

of the numbers of cases within each contour and the corresponding "worst" and "best" case approaches.

Table 41: Calculation of the average probability of loss ranges using data from the contour plot.

Probability Range	Count	Prob. of Loss (Worst Case)	Weighted Prob. of Loss (Worst Case)	Prob. of Loss (Best Case)	Weighted Prob. of Loss (Best Case)
< 0.1	19.83	0.99	0.196	0.90	0.178
0.1 – 0.2	1.37	0.90	0.012	0.80	0.011
0.2 – 0.3	1.56	0.80	0.012	0.70	0.011
0.3 – 0.4	0.72	0.70	0.005	0.60	0.004
0.4 – 0.5	0.48	0.60	0.003	0.50	0.002
0.5 – 0.6	0.61	0.50	0.003	0.40	0.002
0.6 – 0.7	0.61	0.40	0.002	0.30	0.002
0.7 – 0.8	0.97	0.30	0.003	0.20	0.002
0.8 – 0.9	1.10	0.20	0.002	0.10	0.001
> 0.9	72.75	0.10	0.073	0.01	0.007
TOTAL	100	--	0.312	--	0.222

Using the raw data itself, the average probability of loss is calculated to be 0.238. This implies that in 23.8% of cases, the diagnostic program for a 100 component population with the specified input parameters will produce a loss as compared to the "run to failure" program. Comparing this to the values obtained from Table 41, it is clear that the actual probability of loss is within the range of 0.222 to 0.312.

It is also apparent from Table 41 that the greatest risk comes from the lowest probability area of the contour plot. This corresponds to failure rates that are less than 0.03 [failures/component/year]. In these cases, the G/B ratio is at worst 85/15 indicating that the population with G/B ratios greater than 85/15 are basically guaranteed (with probability greater than 90%) to produce a net loss as compared to the "run to failure" program.

The following section will examine the effect of increasing the number of customers on the resulting probabilities of savings.

10.3.1.2 Suburban Region

In switching from the rural region to a suburban region the average number of customers that would be impacted by an outage increases from 20 to 200.

The results for the suburban region are quite similar to those encountered in the rural region. Figure 100 shows the resulting probabilities of savings as a function of failure rate and diagnostic accuracy.

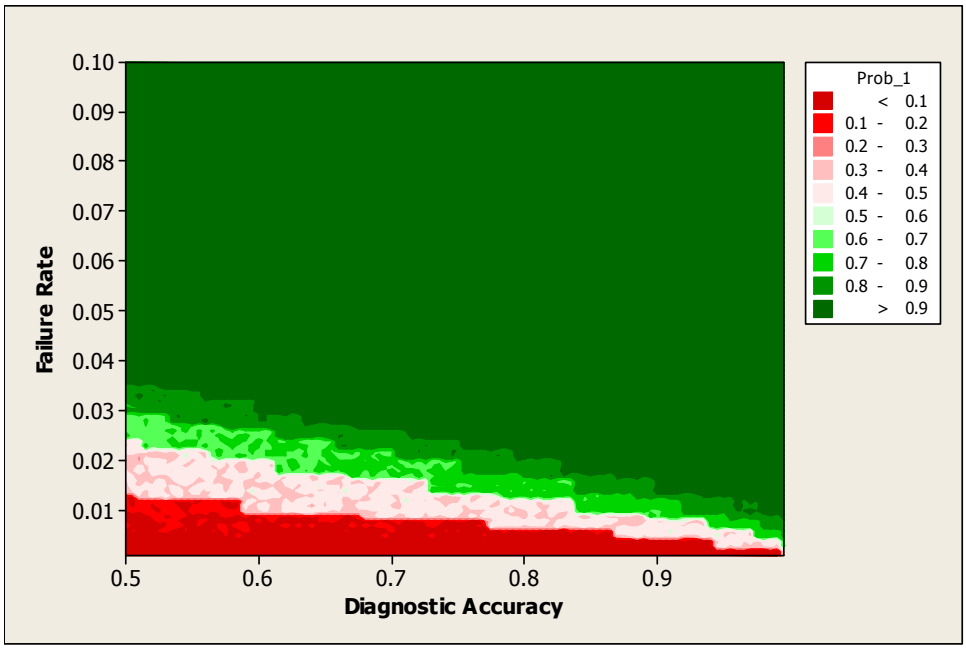


Figure 100: Contour plot of the probability of obtaining savings as a function of failure rate [failures/component/year] and diagnostic accuracy for a suburban region. Note that the jagged surface features are an artifact of the sampling strategy employed and the small number of at-risk components.

Figure 100 shows the same basic structure as was seen in the previous section for the rural case. However, the distribution of the contours is different in that the regions with probabilities less than 90% are reduced in total area as compared to the rural case.

Figure 101 shows the distributions for two different combinations of failure rates and diagnostic accuracy. Note that case two has significant portions of its distribution on the savings side of zero. On the other hand, case one is on the loss side.

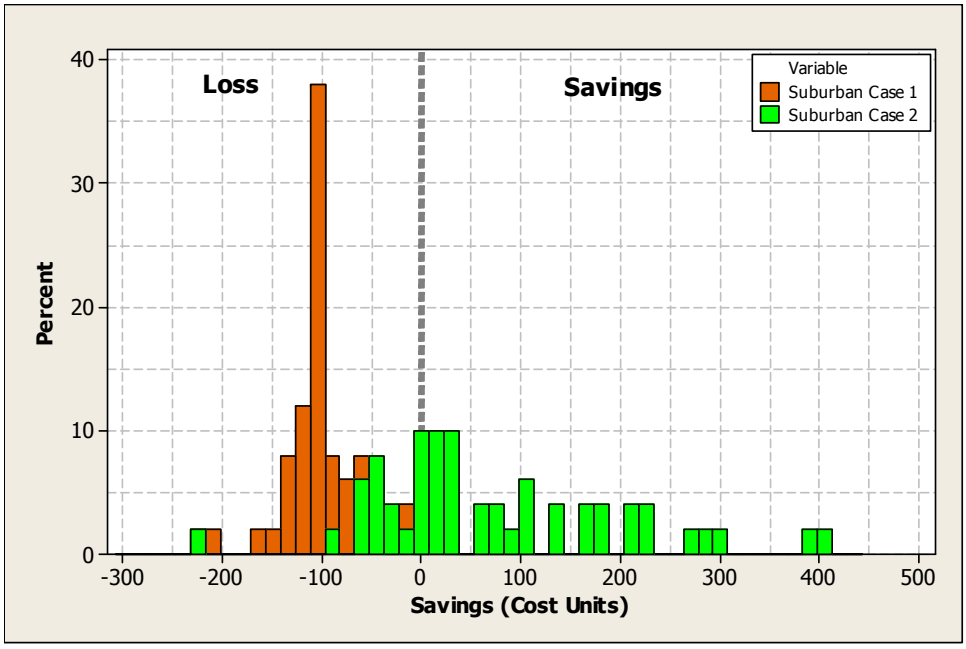


Figure 101: Distributions of savings for two combinations of failure rate and diagnostic accuracy. Case 1: 0.80 accuracy and 0.005 [failures/component/year] and Case 2: 0.80 accuracy and 0.020 [failures/component/year].

Table 42 shows the relative distribution of the area among the different probability intervals.

Table 42: Percentage of data shown in Figure 100 within each of the probability ranges as compared to the rural case.

Probability Range	Rural Count [% of Total Cases]	Suburban Count [% of Total Cases]
< 0.1	19.83	6.05
0.1 – 0.2	1.37	1.21
0.2 – 0.3	1.56	0.03
0.3 – 0.4	0.72	2.44
0.4 – 0.5	0.48	3.93
0.5 – 0.6	0.61	0.13
0.6 – 0.7	0.61	1.92
0.7 – 0.8	0.97	2.79
0.8 – 0.9	1.10	3.90
> 0.9	72.75	77.60

Table 42 shows that the percentage of the cases producing greater than 90% probability of savings increases for the suburban case by almost 5%. Furthermore, the percentage of cases leading to a less than 10% probability of savings reduces from 19.83% to 6.05%, a difference of 13.79%. This represents a reduction of 69.5% as compared to the value found for the rural case.

In the expanded view of the lower third of Figure 100 shown in Figure 102, the difference between the rural and suburban regions is quite apparent.

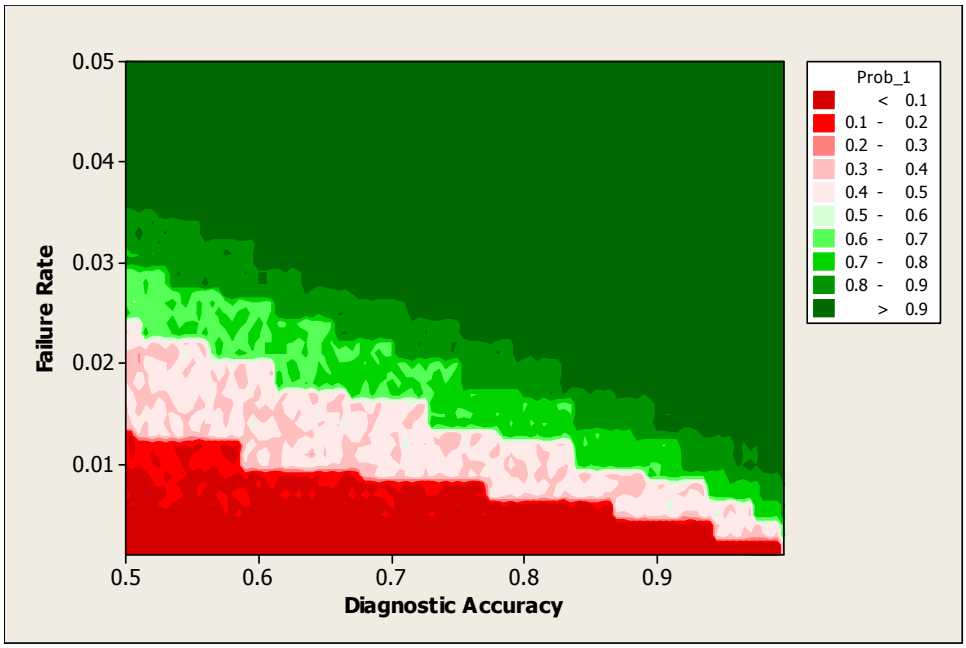


Figure 102: Detailed view of Figure 100 showing the probability contours as a function of failure rate [failures/component/year] and diagnostic accuracy for the suburban region. Note that the jagged surface features are an artifact of the sampling strategy employed.

Based on the data in Table 42, Table 43 shows the resulting ranges of the average probability of loss.

Table 43: Calculation of the average probability of loss ranges using data from the contour plot.

Probability Range	Count [%]	Prob. of Loss (Worst Case)	Weighted Prob. of Loss (Worst Case)	Prob. of Loss (Best Case)	Weighted Prob. of Loss (Best Case)
< 0.1	6.05	0.99	0.060	0.90	0.054
0.1 – 0.2	1.21	0.90	0.011	0.80	0.010
0.2 – 0.3	0.03	0.80	0.000	0.70	0.000
0.3 – 0.4	2.44	0.70	0.017	0.60	0.015
0.4 – 0.5	3.93	0.60	0.024	0.50	0.020
0.5 – 0.6	0.13	0.50	0.001	0.40	0.001
0.6 – 0.7	1.92	0.40	0.008	0.30	0.006
0.7 – 0.8	2.79	0.30	0.008	0.20	0.006
0.8 – 0.9	3.90	0.20	0.008	0.10	0.004
> 0.9	77.60	0.10	0.078	0.01	0.008
TOTAL	100	--	0.214	--	0.122

According to Table 43, the average probability of loss is in the range of 0.122 to 0.214 indicating that there is maximum probability of 21.4% that the diagnostic program would produce a loss. On the other hand, the minimum value at risk is 12.2%. Based on this information, the best a utility could hope for is to achieve an 87.8% probability of generating a savings. Equally, this probability can go as low as 78.6%. Using the raw data, the calculated probability of loss is 0.130 or 13.0%. As with the rural case, this is well within the ranges produced from the contour plots.

10.3.1.3 Urban Region

The urban region case leads to a further reduction in the size of the region with probabilities less than 0.1 of generating a savings. Figure 103 shows the contour plot of the probability as a function of diagnostic accuracy and failure rate. A detailed view of the transition region can be seen in Figure 104.

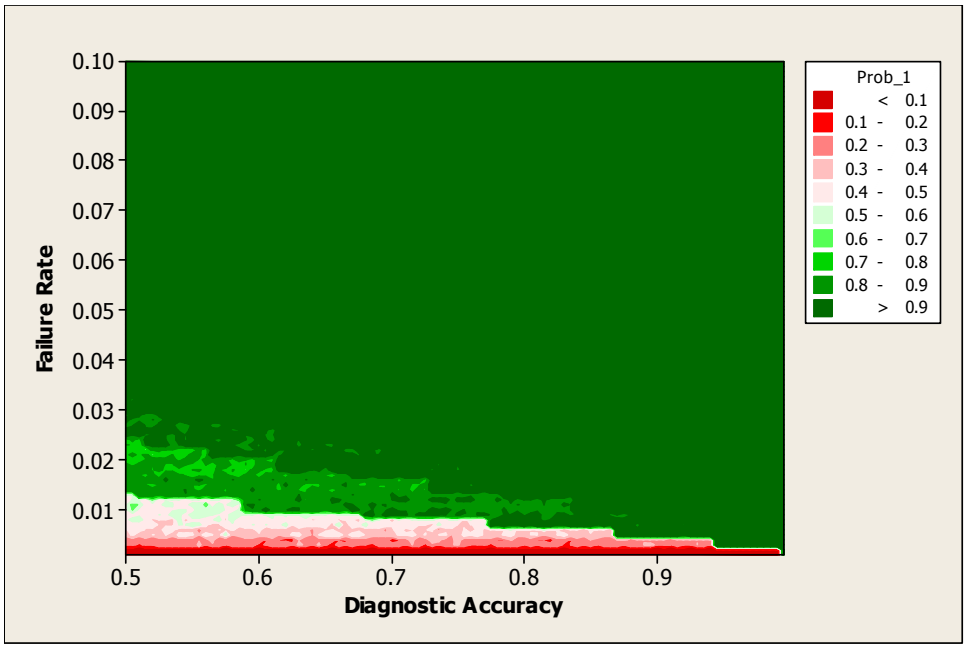


Figure 103: Contour plot of the probability of obtaining savings as a function of failure rate [failures/component/year] and diagnostic accuracy for an urban region.

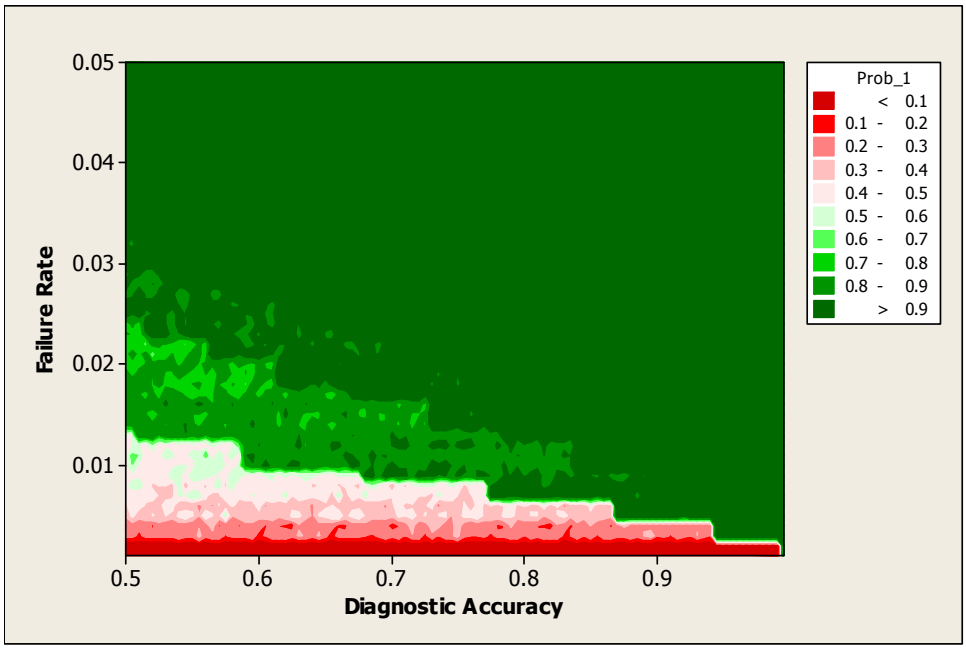


Figure 104: Detailed view of Figure 103 showing the probability contours as a function of failure rate [failures/component/year] and diagnostic accuracy.

Figure 105 shows two distributions resulting from two combinations of failure rates and diagnostic accuracies. Note again that one of the distributions is on the loss side while the other distribution is mainly on the savings side of zero.

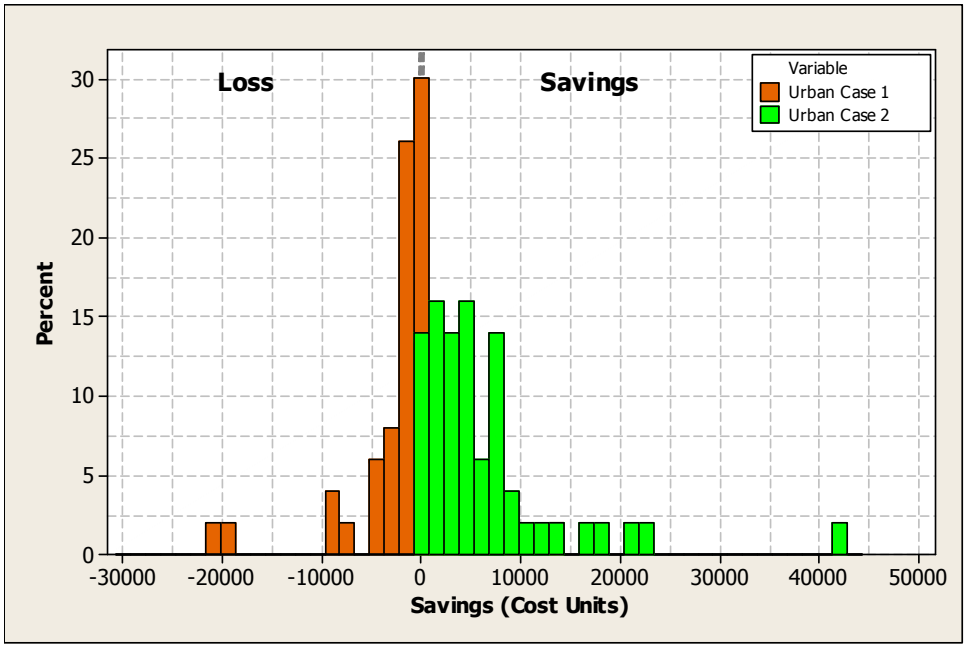


Figure 105: Distributions of savings for two combinations of failure rate and diagnostic accuracy. Case 1: 0.70 accuracy and 0.008 [failures/component/year] and Case 2: 0.70 accuracy and 0.010 [failures/component/year].

The resulting sizes of each of the probability regions are shown in Table 44 along with those for the rural and suburban regions.

Table 44: Percentage of data shown in Figure 100 within each of the probability ranges as compared to the rural case.

Probability Range	Rural Count [% of Total Cases]	Suburban Count [% of Total Cases]	Urban Count [% of Total Cases]
< 0.1	19.83	6.05	1.98
0.1 – 0.2	1.37	1.21	0.26
0.2 – 0.3	1.56	0.03	1.47
0.3 – 0.4	0.72	2.44	1.25
0.4 – 0.5	0.48	3.93	1.66
0.5 – 0.6	0.61	0.13	0.61
0.6 – 0.7	0.61	1.92	0.08
0.7 – 0.8	0.97	2.79	0.87
0.8 – 0.9	1.10	3.90	4.10
> 0.9	72.75	77.60	87.72

Table 44 shows that the urban case produced the smallest percentage of test cases with a probability of less than 0.1 of producing a savings. This is only 10.0% of the rural percentage and 32.7% of the suburban case. Clearly, once the number of customers is high enough the chances of a diagnostic program losing money becomes very small. For the urban region example, only 12.2% of the cases have a probability of savings that is less than 0.90.

Table 45 shows the average or expected value at risk computed using the sizes of the different contour bands.

Table 45: Calculation of the average probability of loss ranges using data from the contour plot shown in Figure 103.

Probability Range	Count [%]	Prob. of Loss (Worst Case)	Weighted Prob. of Loss (Worst Case)	Prob. of Loss (Best Case)	Weighted Prob. of Loss (Best Case)
< 0.1	1.98	0.99	0.020	0.90	0.018
0.1 – 0.2	0.26	0.90	0.002	0.80	0.002
0.2 – 0.3	1.47	0.80	0.012	0.70	0.010
0.3 – 0.4	1.25	0.70	0.009	0.60	0.008
0.4 – 0.5	1.66	0.60	0.010	0.50	0.008
0.5 – 0.6	0.61	0.50	0.003	0.40	0.002
0.6 – 0.7	0.08	0.40	0.000	0.30	0.000
0.7 – 0.8	0.87	0.30	0.003	0.20	0.002
0.8 – 0.9	4.10	0.20	0.008	0.10	0.004
> 0.9	87.72	0.10	0.088	0.01	0.009
TOTAL	100	--	0.154	--	0.063

Computing the expected probability of loss from the raw data produces a value of 0.067 for the set of inputs chosen for this simulation. It appears that using the contours themselves can produce expected ranges that are rather conservative as compared to the results from the full dataset.

10.3.1.4 Comparison of Region Types

Figure 106 shows the each region type's distribution of cases amongst the different contours. As this figure shows, the majority of cases end up in the two extremes with a relatively small percentage of cases lying somewhere in between. This indicates that the transition from the low to high probability groups is very abrupt as compared to the number of cases investigated.

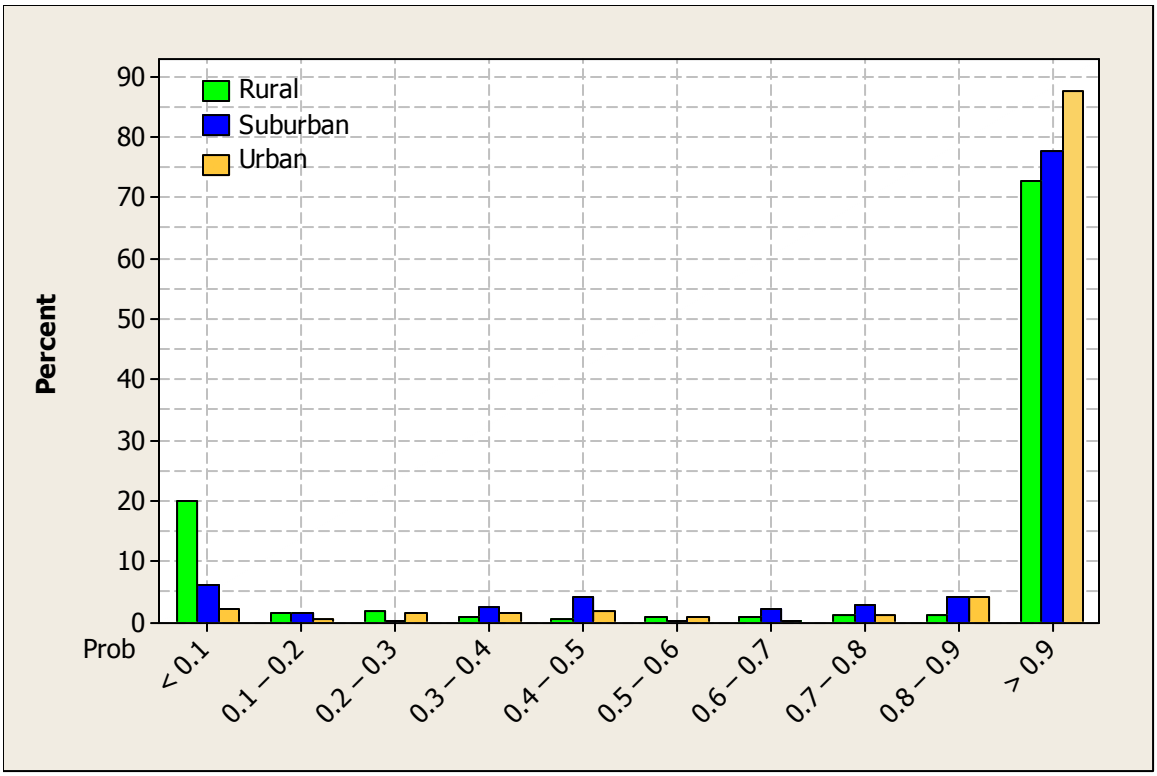


Figure 106: Percentage of samples within each probability range for each region type.

As mentioned in Section 10.3.1.3, Figure 106 shows that increases in the average number of customers leads to a clear reduction in the risk posed by the diagnostic program. One could then argue that a program that makes sense in a rural region will definitely also make sense in an urban region if the only difference between the two is the number of customers.

10.3.2 Customer Penalty Rate and Failure Rate

A similar analysis to the one conducted for the failure rate and diagnostic accuracy can be carried on the customer penalty rate and failure rate. Figure 107, Figure 108, and Figure 109, show the contour plot of the probability of savings as a function of customer penalty rate and failure rate for each region type.

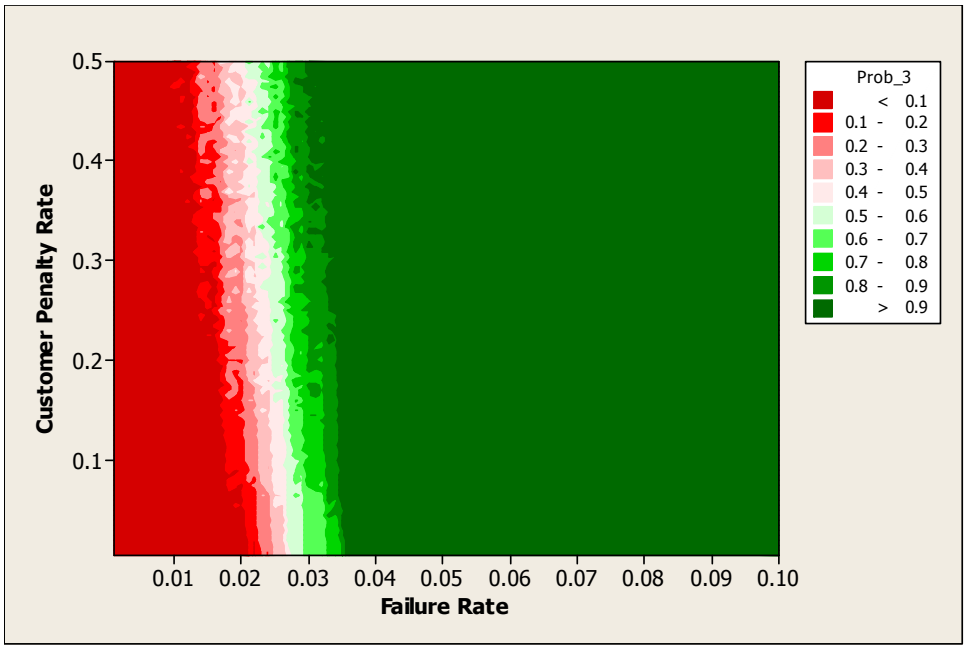


Figure 107: Contour plot of the probability of savings as a function of customer penalty rate [cost units/customer/failure] and failure rate [failures/component/year] for a rural region.

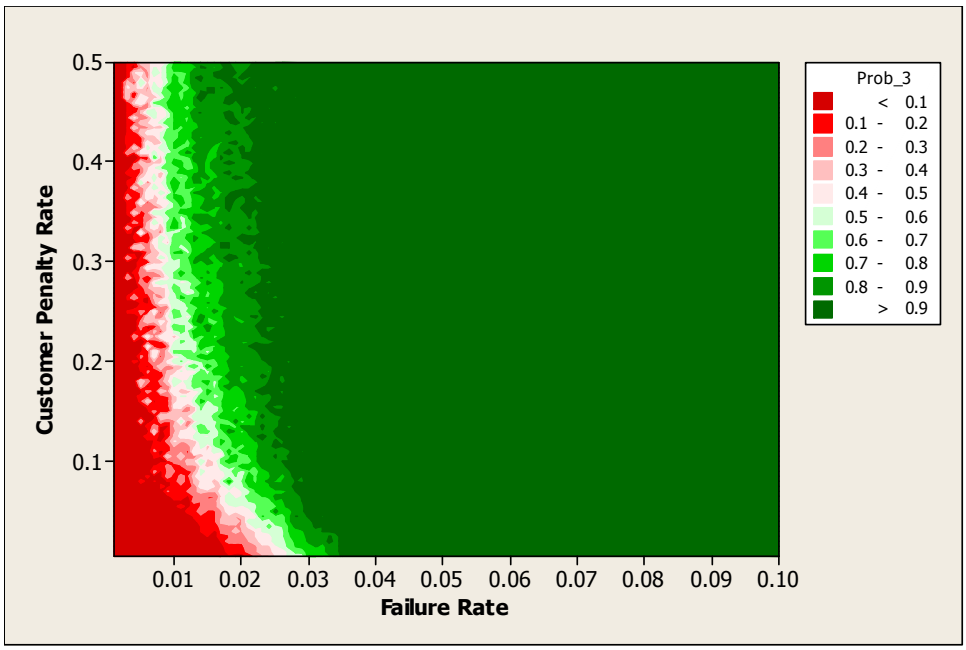


Figure 108: Contour plot of the probability of savings as a function of customer penalty rate [cost units/customer/failure] and failure rate [failures/component/year] for a suburban region.

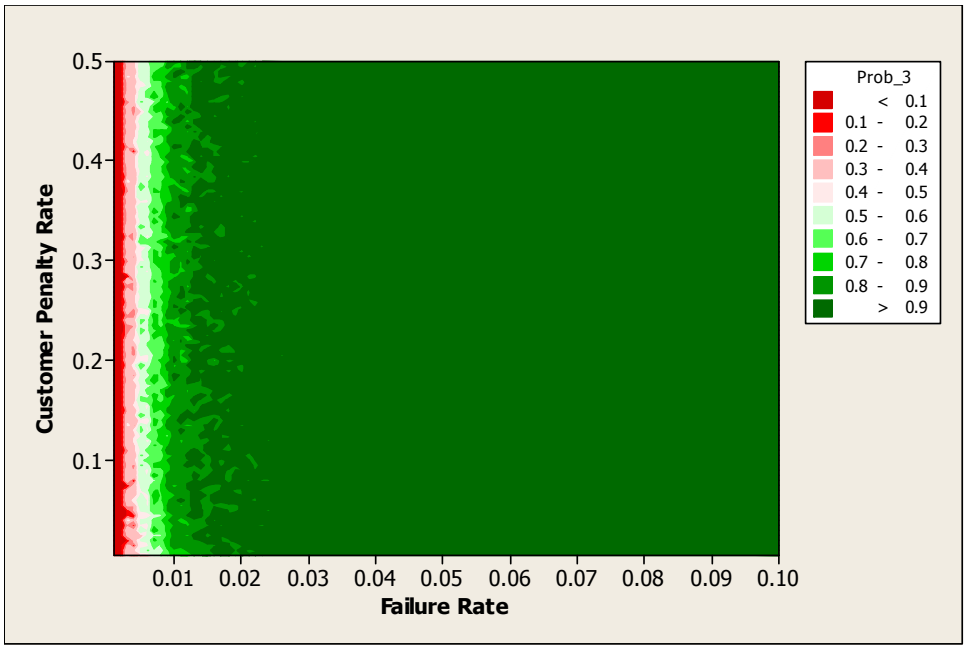


Figure 109: Contour plot of the probability of savings as a function of customer penalty rate [cost units/customer/failure] and failure rate [failures/component/year] for an urban region.

As in Section 10.3.1, the area of the contour representing the low probability cases reduces significantly as the number of customers increases. Furthermore, the customer penalty rate appears to have little effect in the urban case and only minor effect in the other regions. As before, the transition from low to high probability regions is quite sudden and the vast majority of the cases (only cases up to 0.05 [failures/component/year are shown) produce high probabilities of savings (> 0.9). Table 46 shows this numerically.

Table 46: Distribution of the simulated cases among probability ranges.

Probability Range	Rural Count [% of Cases]	Suburban Count [% of Cases]	Urban Count [% of Cases]
< 0.1	15.18	5.28	2.00
0.1 – 0.2	2.38	1.79	0.11
0.2 – 0.3	2.23	1.39	0.36
0.3 – 0.4	2.02	1.43	1.49
0.4 – 0.5	2.01	1.48	0.56
0.5 – 0.6	1.92	1.67	1.35
0.6 – 0.7	1.78	2.08	1.62
0.7 – 0.8	1.97	3.03	1.82
0.8 – 0.9	2.66	4.95	4.75
> 0.9	67.85	76.94	85.94

Table 46 shows the same trend that was observed in Section 10.3.1. As the average number of customers increases, the lower probability contours reduce in size while the higher probability contours increase in size. The identical size at-risk population is much more likely to produce a savings if the program is used in an urban region. This appears to be a general trend.

10.3.3 Customer Penalty Rate and Diagnostic Accuracy

Figure 110, Figure 111, and Figure 112, show the contour plots of the probability of savings as a function of the customer penalty rate and diagnostic accuracy for the rural, suburban, and urban regions, respectively.

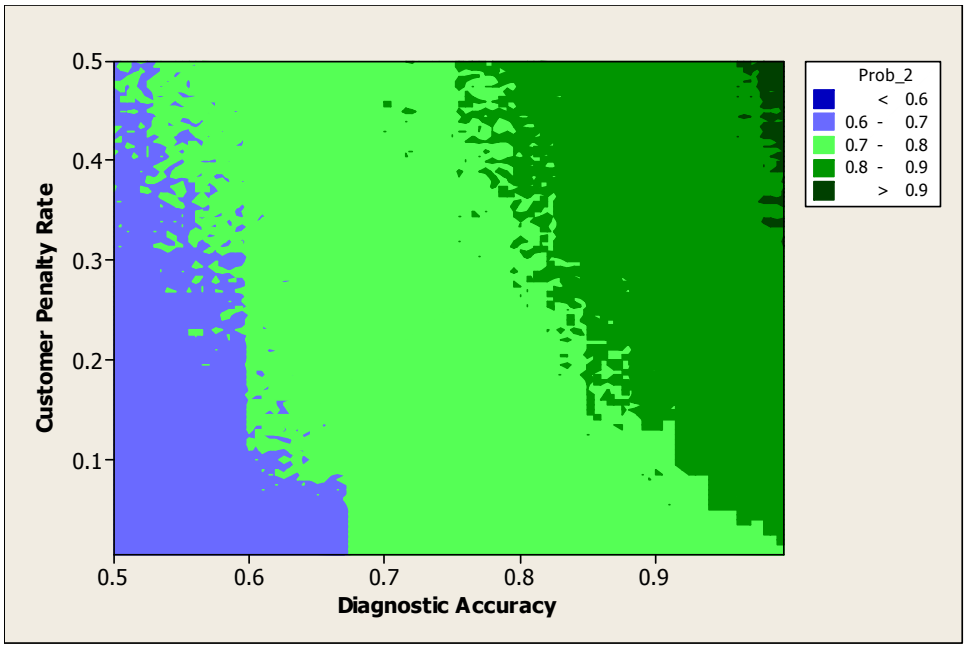


Figure 110: Contour plot of the probability of obtaining savings as a function of the customer penalty rate [cost units/customer/failure] and diagnostic accuracy for a rural region.

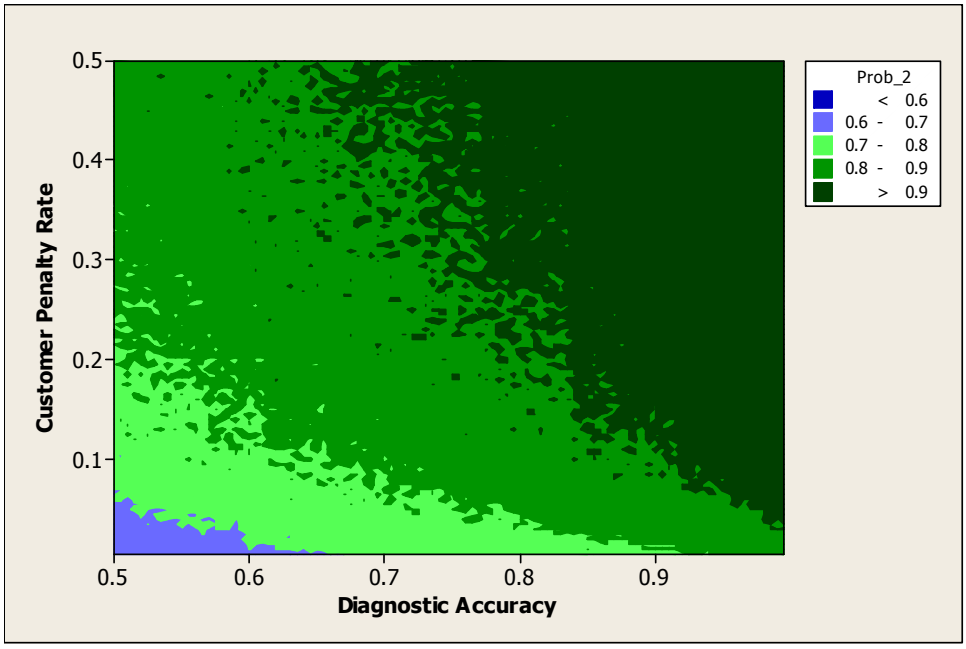


Figure 111: Contour plot of the probability of obtaining savings as a function of the customer penalty rate [cost units/customer/failure] and diagnostic accuracy for a suburban region.

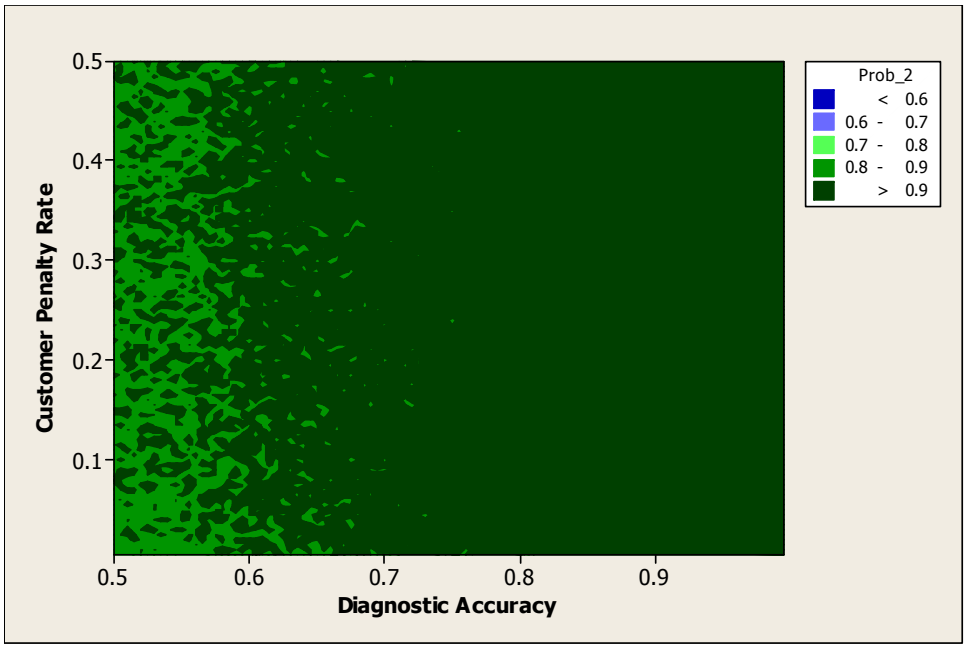


Figure 112: Contour plot of the probability of obtaining savings as a function of the customer penalty rate [cost units/customer/failure] and diagnostic accuracy for an urban region.

As in the other contour plots, as the population increases Figure 112 shows that the area covered by the high probability cases increases. One can compute the sizes of the different contours considering a more detailed scale than is shown in Figure 112. Table 47 shows the percentage of cases that lie within the different contour bands.

Table 47: Distribution of cases amongst the different contour bands.

Probability Range	Rural Count [% of Cases]	Suburban Count [% of Cases]	Urban Count [% of Cases]
< 0.6	0.00	0.00	0.00
0.6 – 0.7	17.38	1.68	0.00
0.7 – 0.8	50.67	12.39	0.00
0.8 – 0.9	30.71	45.37	13.07
> 0.9	1.24	40.56	86.93

Table 47 again illustrates the differences in the region types only the shift from one probability band to the next is more apparent.

10.3.4 Value at risk and Probability of Loss

This section summarizes the value at risk and probability that each of the cases presented thus far should be expected to produce. Table 48 shows the value at risk for each of the region types for a confidence level of 95%.

Table 48: Value at risk for each region type for 95% confidence.

Region Type	Value at Risk [Cost Units]
Rural	116
Suburban	160
Urban	5864

Note that the smallest value at risk occurs in the rural region. In this case, the value at risk is only 116 cost units as compared to more than 5800 in the urban case. This information can be combined with the probability of loss as shown in Table 49.

Table 49: Expected probability of loss for each region type.

Region Type	Expected Prob. of Loss
Rural	0.238
Suburban	0.130
Urban	0.067

Table 48 and Table 49 show a trend that will likely be common to the larger at-risk population example in the next section. These tables combined show that the risk of the

diagnostic program creating a loss is smallest for the urban regions. However, that region type also has the highest value at risk. In other words, the utility might lose more often in the rural regions but when it does so the amount of loss incurred is small as compared to an urban region. On the other hand, in the urban region losses will occur less frequently but they will be much more expensive than the rural region when they do occur.

10.4 MEDIUM SIZE AT-RISK POPULATION (1000 COMPONENTS)

The analysis and results in this section will follow the same pattern as in Section 10.3. The probabilities of obtaining savings from the program (as compared to a "run to failure" program) for a 1000 component at-risk population will be presented as functions of diagnostic accuracy, failure rate, and customer penalty cost. As before, rural, suburban, and urban, regions are investigated.

10.4.1 Failure Rate and Diagnostic Accuracy

As in Section 10.4.1, this section will review the behavior of the model in terms of the diagnostic accuracy and failure rate.

10.4.1.1 Rural Region

Figure 113 shows a contour plot of the probability of achieving savings as a function of the failure rate and diagnostic accuracy for a rural region.

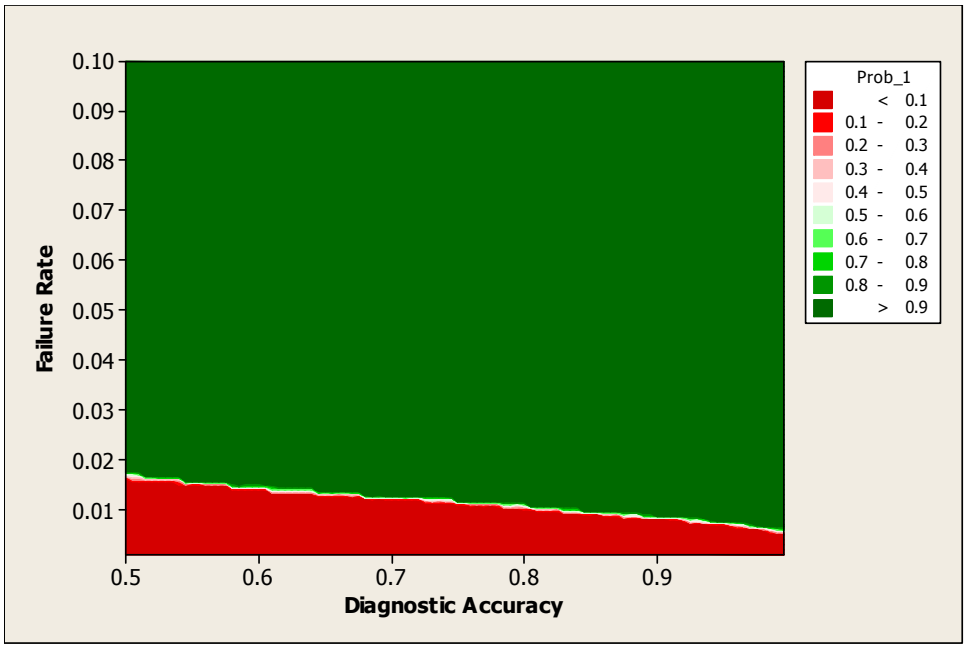


Figure 113: Contour plot of the probability of obtaining savings as a function of the failure rate [failures/component/year] and diagnostic accuracy for a rural region.

As in Section 10.3.1.1, Figure 113 clearly shows regions where the probability of obtaining a cost savings as compared to "run to failure" is less than 20%. As before, there are failure rates for which even a 100% accurate diagnostic technique will not yield a savings. Unlike the 100 component case, this only affects the cases with failure rates that are less than 0.02 [failures/component/year] (or 20 [failures/1000 components/year]). For failure rates greater than this any diagnostic with an accuracy of 51% or more is able to produce a savings with at least 90% probability. Within the five year horizon of this simulation, the total number of expected failures at this failure rate is approximately 100 failures. This translates to a G/B ratio of 90/10.

By utilizing a more accurate diagnostic, the minimum failure rate needed to guarantee a 90% probability of a net program savings is less than that required for the lower accuracies. Figure 114 shows a close up view of the transition between low probabilities

(less than 0.3) to high probabilities of savings (greater than 0.7). This transition is far more abrupt than was seen in the 100 component case.

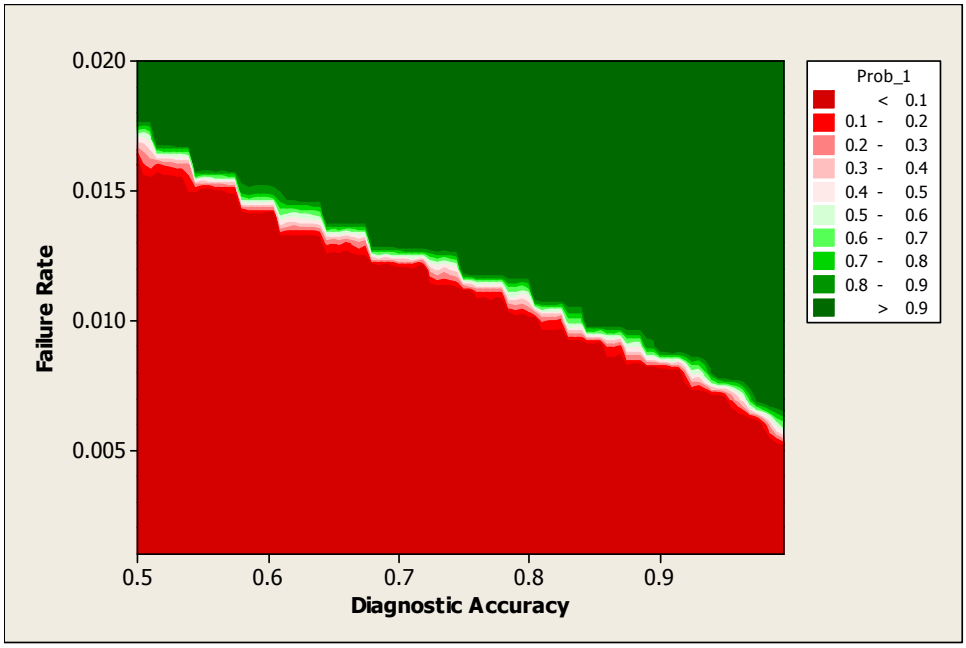


Figure 114: Detailed view of Figure 113 showing the probability contours as a function of failure rate [failures/component/year] and diagnostic accuracy.

According to Figure 114, the probability of savings of 90% can be obtained with a failure rate of approximately 0.008 [failures/component/year] (8 [failures/1000 components/year]) and a diagnostic accuracy greater than 95%. This failure rate is 60% less than the failure rate required for a diagnostic accuracy of 51%. As a result, one could utilize an increased G/B ratio of 96/4 with the more accurate diagnostic techniques while still achieving the same probability of savings.

Figure 115 shows the full distributions for two combinations of failure rates and diagnostic accuracies. Clearly, portions of the case two distribution are in the savings region while case one is primarily on the loss side of zero.

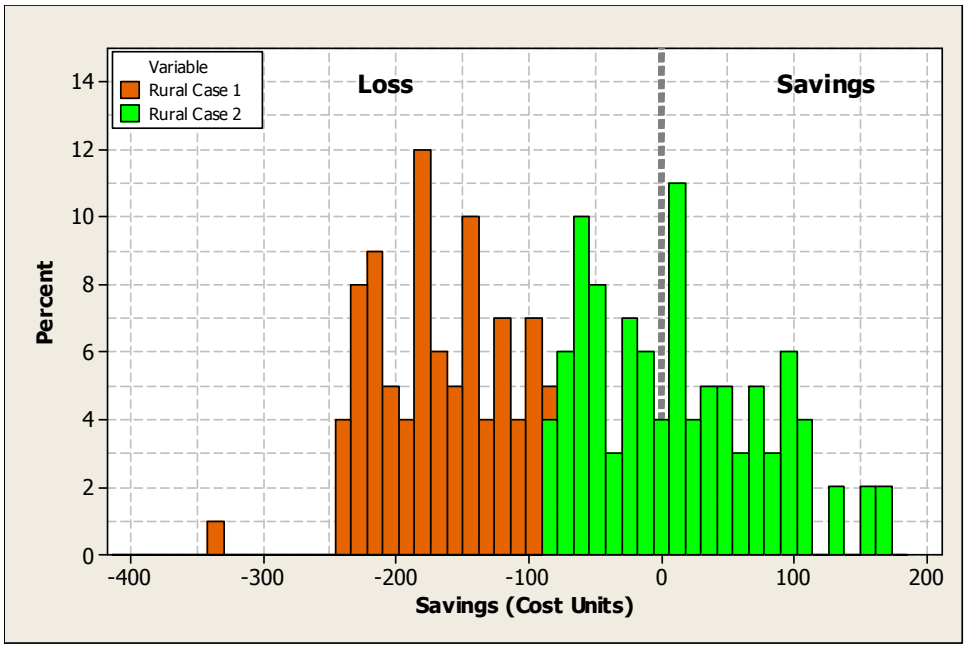


Figure 115: Distributions of savings for two combinations of failure rate and diagnostic accuracy. Case 1: 0.80 accuracy and 0.010 [failures/component/year] and Case 2: 0.80 accuracy and 0.011 [failures/component/year].

Table 50 shows the resulting percentages of cases that lie within each probability range (contour).

Table 50: Percentage of data shown in Figure 113 within each of the probability ranges.

Probability Range	Count [% of Total Cases]
< 0.1	10.66
0.1 – 0.2	0.18
0.2 – 0.3	0.16
0.3 – 0.4	0.04
0.4 – 0.5	0.11
0.5 – 0.6	0.13
0.6 – 0.7	0.05
0.7 – 0.8	0.01
0.8 – 0.9	0.09
> 0.9	88.57

As Table 50 shows, the vast majority of cases shown in Figure 113 (88.6%) have a probability of savings that is greater than 90%. This is again the result of the G/B ratio and is 16% higher than the 100 component case. Also different from the previous case is the reduction from 19.8% to 10.7% the percentage of cases with less than 10% probability of savings. This difference is due to the reduction in failure rate from 0.04 to 0.02 [failures/component/year] needed to guarantee a 90% probability of savings.

The probability of loss may also be calculated from the contour plots as shown in Table 51.

Table 51: Calculation of the average probability of loss ranges using data from the contour plot.

Probability Range	Count	Prob. of Loss (Worst Case)	Weighted Prob. of Loss (Worst Case)	Prob. of Loss (Best Case)	Weighted Prob. of Loss (Best Case)
< 0.1	10.66	0.99	0.106	0.90	0.096
0.1 – 0.2	0.18	0.90	0.002	0.80	0.001
0.2 – 0.3	0.16	0.80	0.001	0.70	0.001
0.3 – 0.4	0.04	0.70	0.000	0.60	0.000
0.4 – 0.5	0.11	0.60	0.001	0.50	0.001
0.5 – 0.6	0.13	0.50	0.001	0.40	0.001
0.6 – 0.7	0.05	0.40	0.000	0.30	0.000
0.7 – 0.8	0.01	0.30	0.000	0.20	0.000
0.8 – 0.9	0.09	0.20	0.000	0.10	0.000
> 0.9	88.57	0.10	0.089	0.01	0.009
TOTAL	100	--	0.199	--	0.109

Using the raw data itself, the average probability of loss is calculated to be 0.111. This implies that in 11.1% of cases, the diagnostic program for a 1000 component population will lead to a loss. Comparing this to the values obtained from Table 51, it is clear that the actual probability of loss is again within the range of 0.109 to 0.199.

As in the case of the 100 component scenario, the risk is essentially split between the two extremes resulting from probabilities less than 10% and greater than 90%. As

mentioned, the transition between these extremes is rather abrupt with only about 0.8% the cases lying in the transition region. This means that the diagnostic program will either yield a savings or it clearly will not. There is no middle ground based on the assumptions about uncertainties used in the example.

As with the 100 component case, it is also valuable to examine the effect of increasing the number of customers on the resulting probabilities of savings.

10.4.1.2 Suburban Region

The increase in customers from 20 to 200 leads to the contour plot shown in Figure 116 for diagnostic accuracy and failure rate.

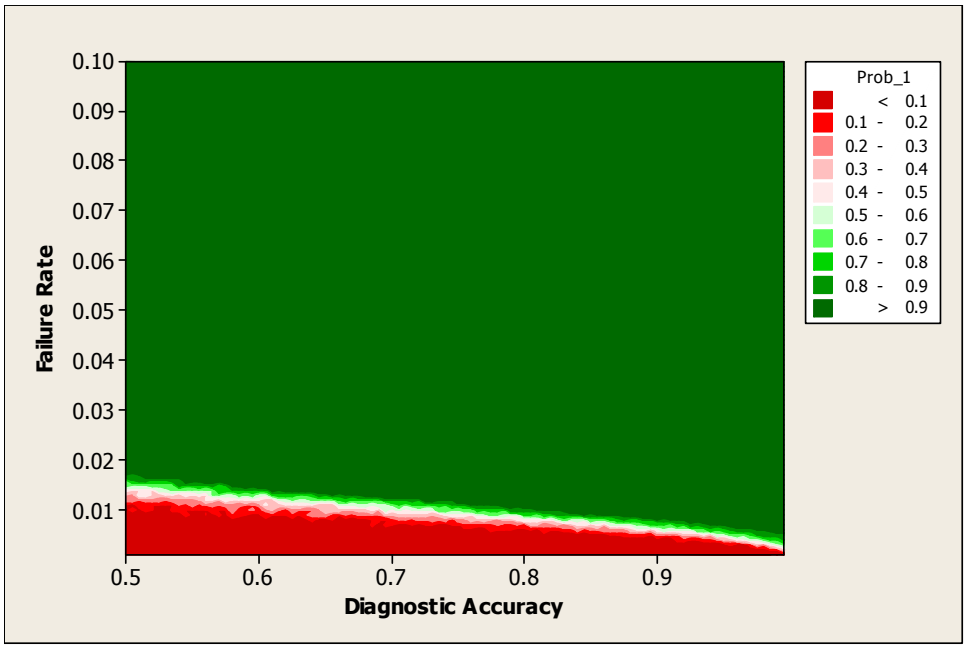


Figure 116: Contour plot of the probability of obtaining savings as a function of failure rate [failures/component/year] and diagnostic accuracy for a suburban region.

Figure 116 shows once more the same structure that has been seen in all previous cases. For failure rates that are greater than 0.02 [failures/component/year], all the cases

have a 90% probability of producing a savings. This boundary can be extended to lower failure rates by increasing the diagnostic accuracy as indicated by the negative slope of the red contour in Figure 116. However, this effect is relatively small.

Figure 115 Figure 117 shows the full distributions for two combinations of failure rates and diagnostic accuracies. Again, the majority of the case two distribution is in the savings region while case one is split between the loss and savings sides.

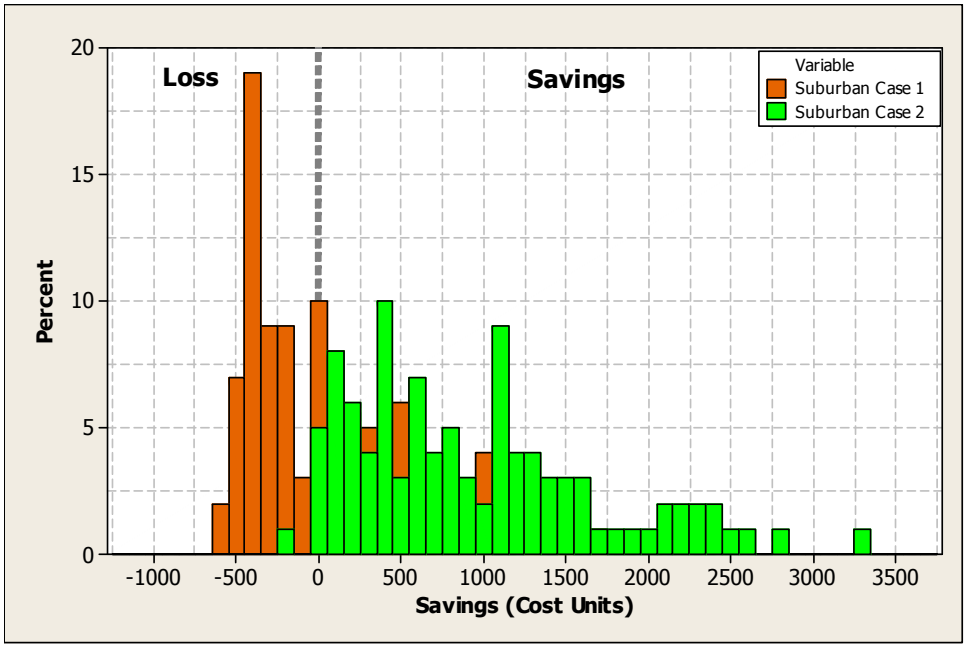


Figure 117: Distributions of savings for two combinations of failure rate and diagnostic accuracy. Case 1: 0.70 accuracy and 0.010 [failures/component/year] and Case 2: 0.70 accuracy and 0.013 [failures/component/year].

Table 52 shows the relative distribution of the area amongst the different probability intervals.

Table 52: Percentage of data shown in Figure 116 within each of the probability ranges as compared to the rural case.

Probability Range	Rural Count [% of Total Cases]	Suburban Count [% of Total Cases]
< 0.1	10.66	6.17
0.1 – 0.2	0.18	0.92
0.2 – 0.3	0.16	0.67
0.3 – 0.4	0.04	0.58
0.4 – 0.5	0.11	0.40
0.5 – 0.6	0.13	0.53
0.6 – 0.7	0.05	0.52
0.7 – 0.8	0.01	0.52
0.8 – 0.9	0.09	0.56
> 0.9	88.57	89.13

Table 52 shows s slight increase in the greater than 0.90 category but the biggest difference is the reduction in the less than 0.1 probability category. In the suburban case, this category now accounts for 4.5% less of the cases. This amounts to a reduction of 42.1% from the rural case. These cases are then distributed amongst the transition contour regions. As a result, the transition from the low to high probability regions is somewhat smoother than the rural case. This is shown graphically in Figure 118.

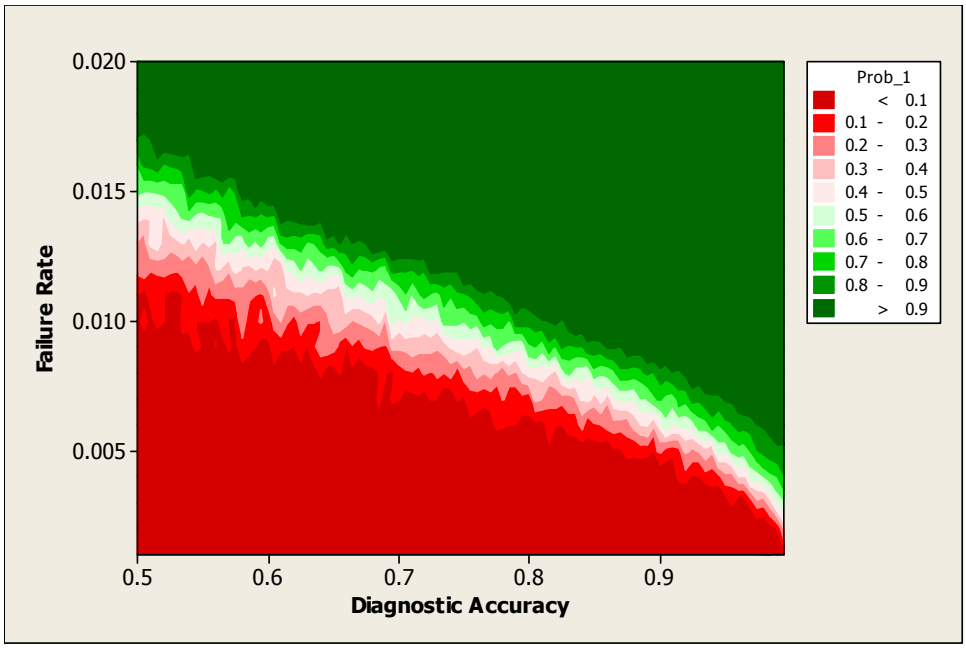


Figure 118: Detailed view of Figure 116 showing the probability contours as a function of failure rate [failures/component/year] and diagnostic accuracy for the suburban region.

Table 53 shows the resulting probability of loss calculation for the contours shown in Figure 116 and Figure 118.

Table 53: Calculation of the average probability of loss ranges using data from the contour plot.

Probability Range	Count	Prob. of Loss (Worst Case)	Weighted Prob. of Loss (Worst Case)	Prob. of Loss (Best Case)	Weighted Prob. of Loss (Best Case)
< 0.1	6.17	0.99	0.061	0.90	0.056
0.1 – 0.2	0.92	0.90	0.008	0.80	0.007
0.2 – 0.3	0.67	0.80	0.005	0.70	0.005
0.3 – 0.4	0.58	0.70	0.004	0.60	0.003
0.4 – 0.5	0.40	0.60	0.002	0.50	0.002
0.5 – 0.6	0.53	0.50	0.003	0.40	0.002
0.6 – 0.7	0.52	0.40	0.002	0.30	0.002
0.7 – 0.8	0.52	0.30	0.002	0.20	0.001
0.8 – 0.9	0.56	0.20	0.001	0.10	0.001
> 0.9	89.13	0.10	0.089	0.01	0.009
TOTAL	100	--	0.178	--	0.087

Table 53 shows a range of 0.087 to 0.178 for the average probability of loss while the true value assuming all cases have equal probability is 0.087. As in the other cases, this value is within the range shown in Table 53 although at one of the extremes.

10.4.1.3 Urban Region

As in the previous cases, Figure 119 and Figure 120 show the probability contour plots as a function of diagnostic accuracy and failure rate.

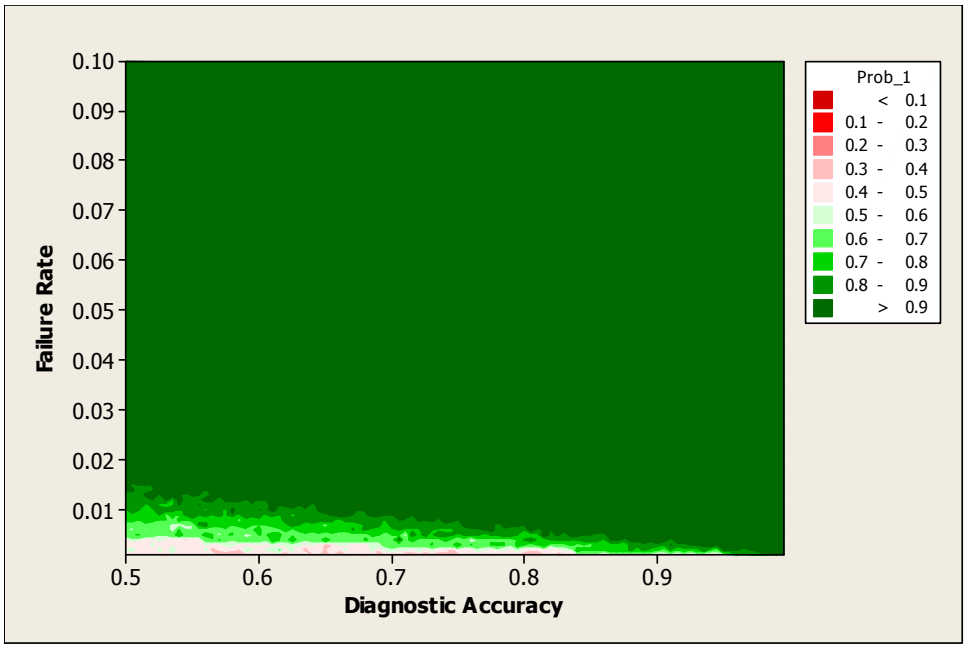


Figure 119: Contour plot of the probability of obtaining savings as a function of failure rate [failures/component/year] and diagnostic accuracy for an urban region.

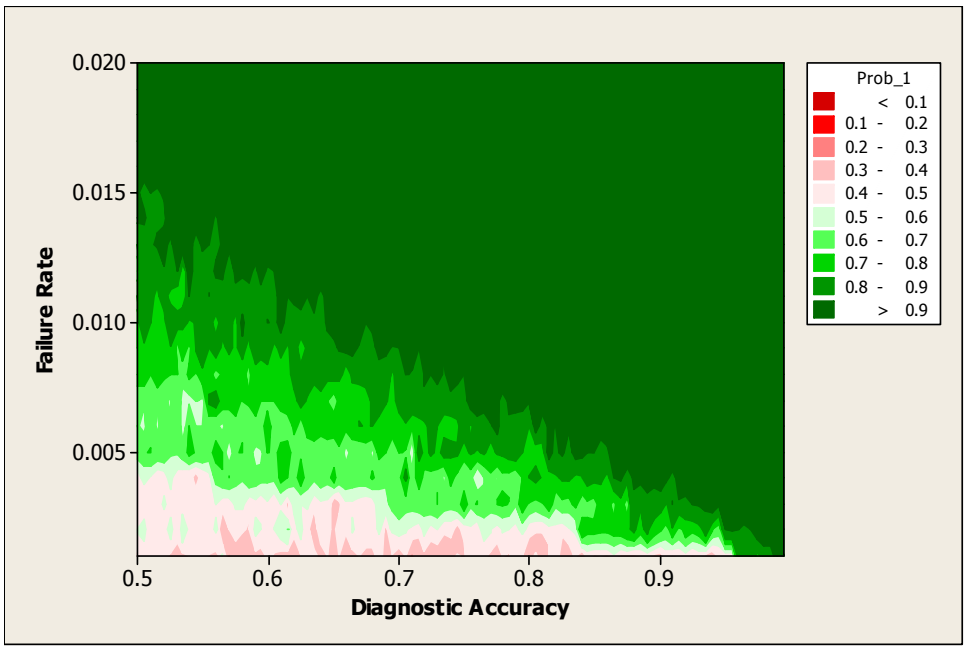


Figure 120: Detailed view of Figure 119 showing the probability contours as a function of failure rate [failures/component/year] and diagnostic accuracy.

Figure 115 Figure 117 shows the full distributions for two combinations of failure rates and diagnostic accuracies. In both cases, the distributions appear to be split relatively evenly between the loss and savings sides.

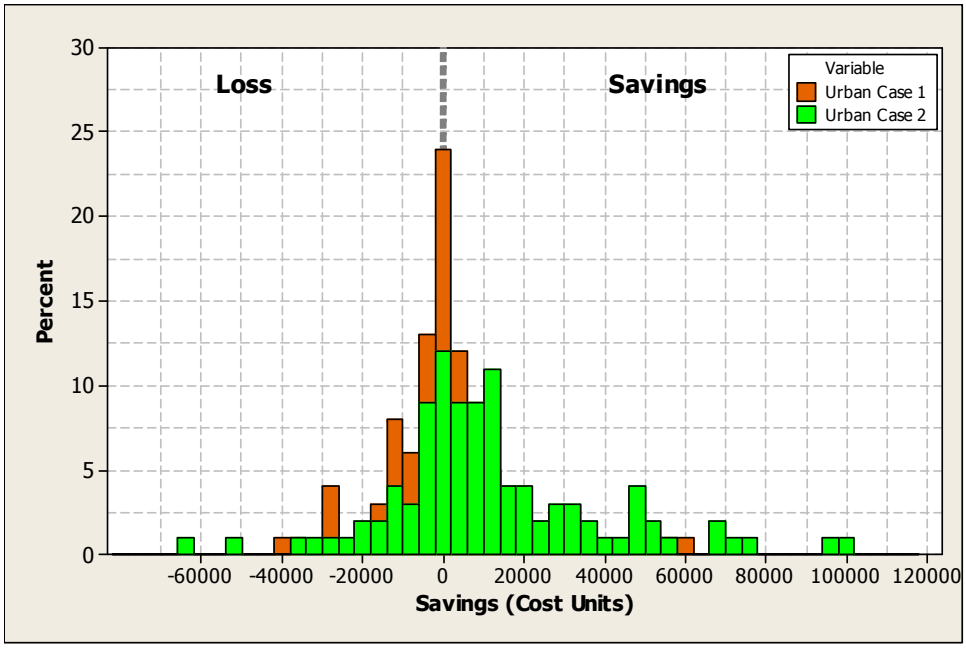


Figure 121: Distributions of savings for two combinations of failure rate and diagnostic accuracy. Case 1: 0.70 accuracy and 0.002 [failures/component/year] and Case 2: 0.70 accuracy and 0.004 [failures/component/year].

Table 54 shows the distribution of cases amongst the different contours shown in Figure 119.

Table 54: Percentage of data shown in Figure 119 within each of the probability ranges as compared to the rural and suburban cases.

Probability Range	Rural Count [% of Total Cases]	Suburban Count [% of Total Cases]	Urban Count [% of Total Cases]
< 0.1	10.66	6.17	0.00
0.1 – 0.2	0.18	0.92	0.00
0.2 – 0.3	0.16	0.67	0.04
0.3 – 0.4	0.04	0.58	0.60
0.4 – 0.5	0.11	0.40	1.10
0.5 – 0.6	0.13	0.53	0.53
0.6 – 0.7	0.05	0.52	1.24
0.7 – 0.8	0.01	0.52	1.55
0.8 – 0.9	0.09	0.56	1.73
> 0.9	88.57	89.13	93.21

Table 55 shows the resulting probability of loss calculation for the contours shown in Figure 119 and Figure 120.

Table 55: Calculation of the average probability of loss ranges using data from the contour plot.

Probability Range	Count	Prob. of Loss (Worst Case)	Weighted Prob. of Loss (Worst Case)	Prob. of Loss (Best Case)	Weighted Prob. of Loss (Best Case)
< 0.1	0.00	0.99	0.000	0.90	0.000
0.1 – 0.2	0.00	0.90	0.000	0.80	0.000
0.2 – 0.3	0.04	0.80	0.000	0.70	0.000
0.3 – 0.4	0.60	0.70	0.004	0.60	0.004
0.4 – 0.5	1.10	0.60	0.007	0.50	0.006
0.5 – 0.6	0.53	0.50	0.003	0.40	0.002
0.6 – 0.7	1.24	0.40	0.005	0.30	0.004
0.7 – 0.8	1.55	0.30	0.005	0.20	0.003
0.8 – 0.9	1.73	0.20	0.003	0.10	0.002
> 0.9	93.21	0.10	0.093	0.01	0.009
TOTAL	100	--	0.120	--	0.029

Table 55 shows a range of 0.029 to 0.120 for the average probability of loss while the true value assuming all cases have equal probability is 0.026. In this case, the true probability of loss is outside the range shown in Table 55.

10.4.1.4 Comparison of Region Types

Figure 122 shows distribution of cases for each region type. Note that very few cases are located between the two extremes for each of the region types.

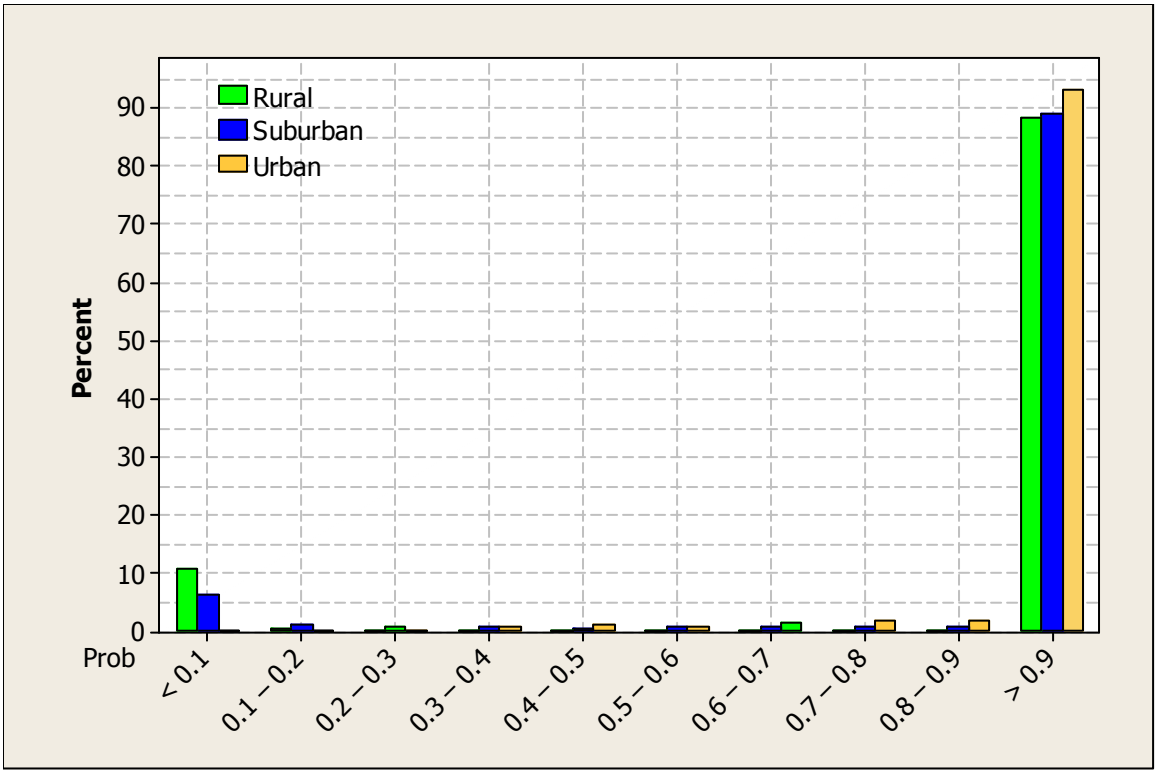


Figure 122: Percentage of samples within each probability range for each region type for the 1000 component at-risk population.

Clearly, once the number of potential impacted customers reaches a certain level, the probability of loss posed by the diagnostic program becomes quite low regardless of the choice of diagnostic. In these cases, the less expensive diagnostic tests may be preferred even though they might deliver less accuracy than their more expensive counterparts. On

the other hand, the diagnostic accuracy is a more important factor where the G/B ratio is rather high (greater than 85/15).

10.4.2 Customer Penalty Rate and Failure Rate

As in the 100 component population case, the probabilities of savings may be examined from the perspective of customer penalty rate and failure rate. Figure 123, Figure 124, and Figure 125, show the resulting contour plots for the rural, suburban, and urban cases, respectively.

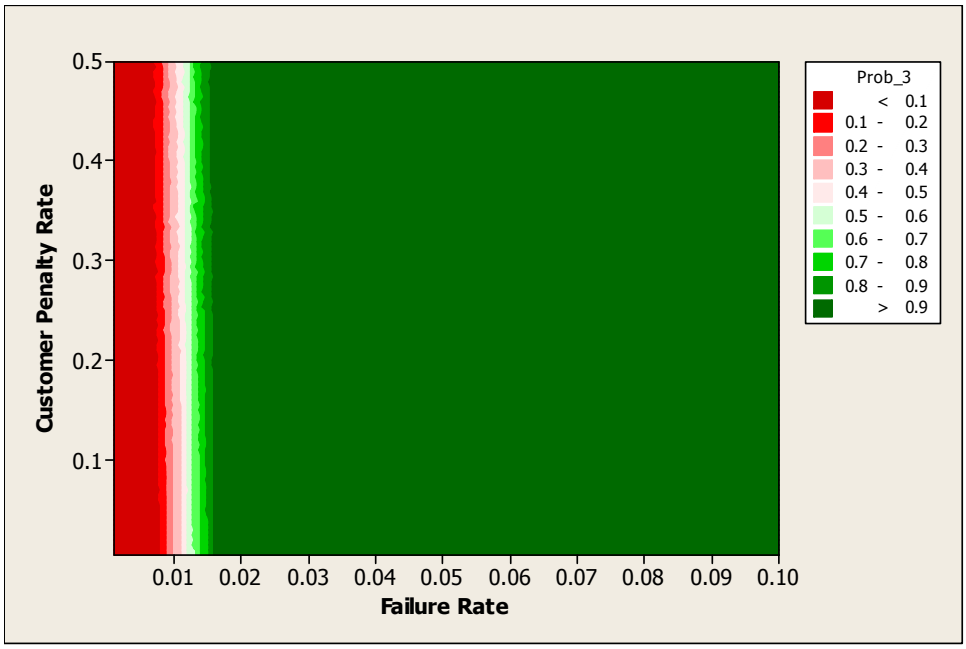


Figure 123: Contour plot of the probability of obtaining savings as a function of the customer penalty rate [cost units/customer/failure] and component failure rate [failures/component/year] for a rural region.

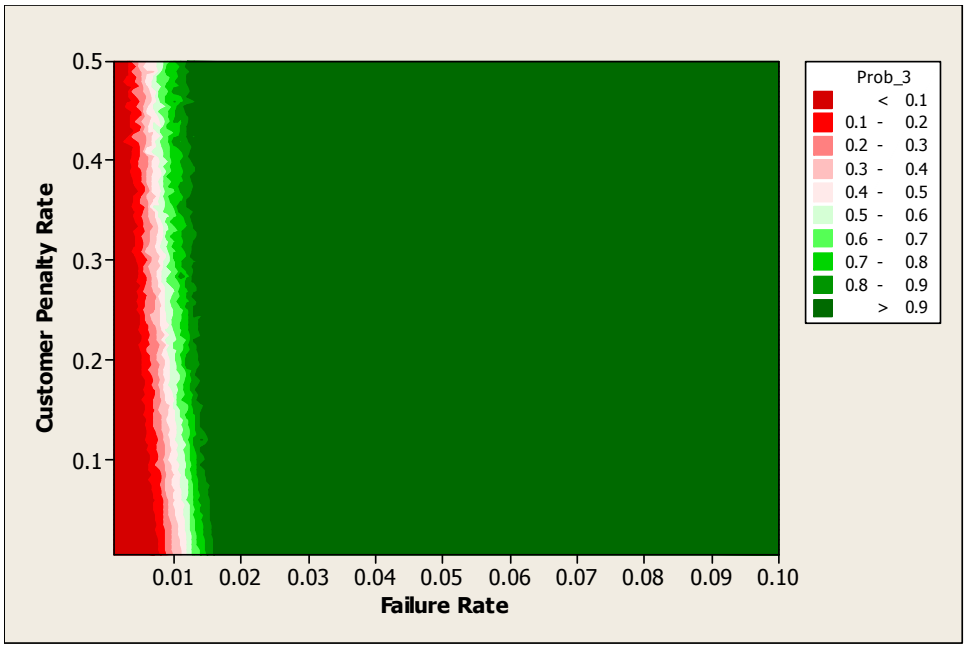


Figure 124: Contour plot of the probability of obtaining savings as a function of the customer penalty rate [cost units/customer/failure] and component failure rate [failures/component/year] for a suburban region.

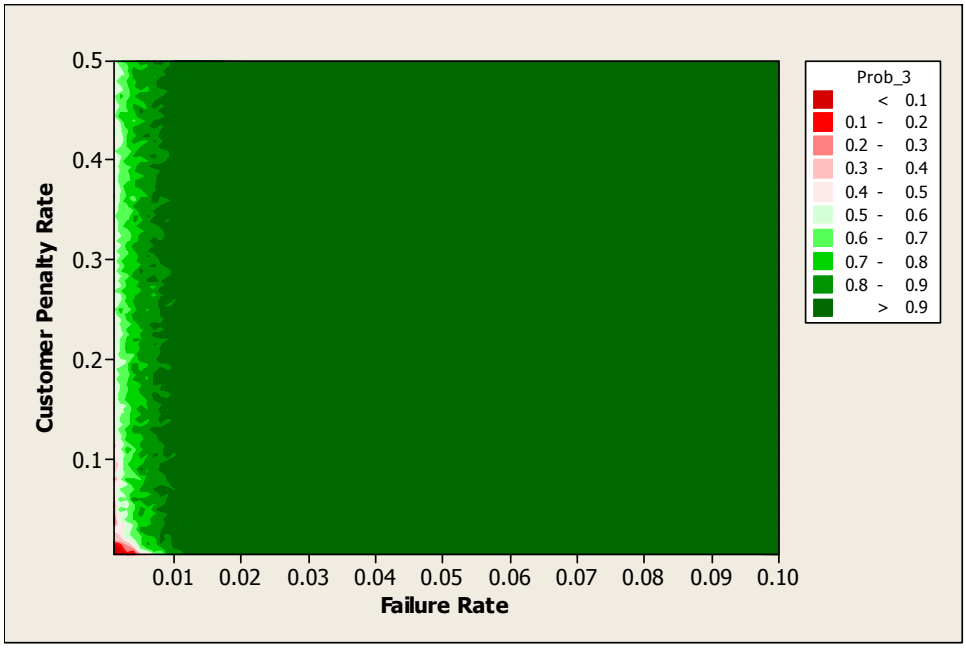


Figure 125: Contour plot of the probability of obtaining savings as a function of the customer penalty rate [cost units/customer/failure] and component failure rate [failures/component/year] for an urban region.

The above figure shows that as the number of customers impacted increases, the number of cases yielding probabilities of savings that are less than 0.1 decreases. This is evidenced by the reduction in the size of the red contour. Furthermore, it is apparent that for customer penalty rates in the range of [0.01, 0.5] the penalty itself has little effect on the likelihood of a successful program. In terms of the actual cost of the component and installation, the penalty per customer represents a maximum of 25% of this cost. That is a relatively large percentage considering that even in the rural case the average number of customers served by each component is 20. Therefore, it is surprising that the customer penalty cost does not exert more influence over the probability of savings.

One explanation for the limited effect of the customer penalty is the fact that the penalty affects both the diagnostic program cost as well as the base case program ("run to failure"). Some of the influence is likely being cancelled out by the presence of these costs in both programs.

Table 56 shows the distribution of the cases among the different contour levels.

Table 56: Distribution of the simulated cases amongst probability ranges.

Probability Range	Rural Count [% of Cases]	Suburban Count [% of Cases]	Urban Count [% of Cases]
< 0.1	6.93	4.42	0.09
0.1 – 0.2	1.07	1.28	0.06
0.2 – 0.3	1.00	0.97	0.07
0.3 – 0.4	1.32	0.95	0.15
0.4 – 0.5	0.69	0.94	0.51
0.5 – 0.6	1.01	0.89	0.87
0.6 – 0.7	0.93	1.01	1.18
0.7 – 0.8	1.09	1.18	1.66
0.8 – 0.9	0.96	1.26	2.94
> 0.9	85.00	87.10	92.47

The observed shift in the sizes of the probability bands for the 100 component population also holds for the 1000 component at-risk population. The difference, though, between these two scenarios is not as large as in the 100 component case.

10.4.3 Customer Penalty Rate and Diagnostic Accuracy

Figure 126, Figure 127, and Figure 128, show the contour plots of the probability of savings as a function of the customer penalty rate and diagnostic accuracy for each of the region types.

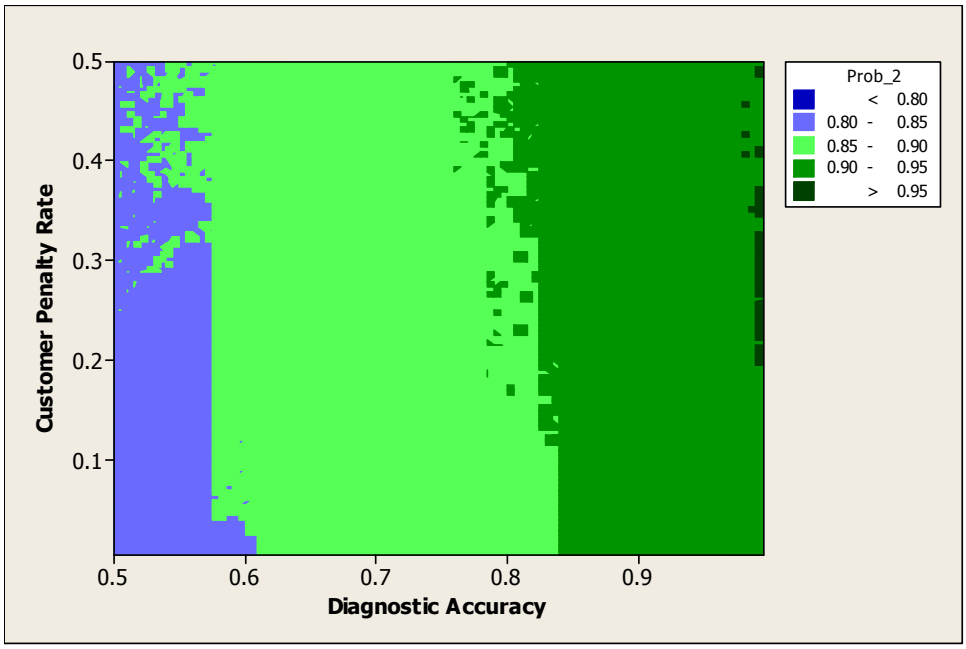


Figure 126: Contour plot of the probability of obtaining savings as a function of the customer penalty rate [cost units/customer/failure] and diagnostic accuracy for a rural region.

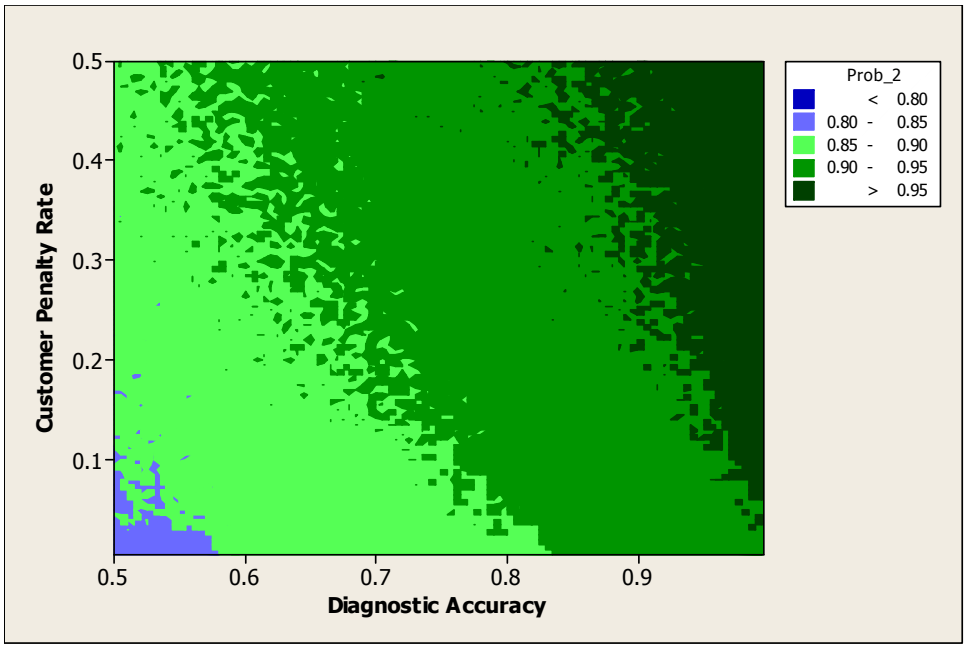


Figure 127: Contour plot of the probability of obtaining savings as a function of the customer penalty rate [cost units/customer/failure] and diagnostic accuracy for a suburban region.

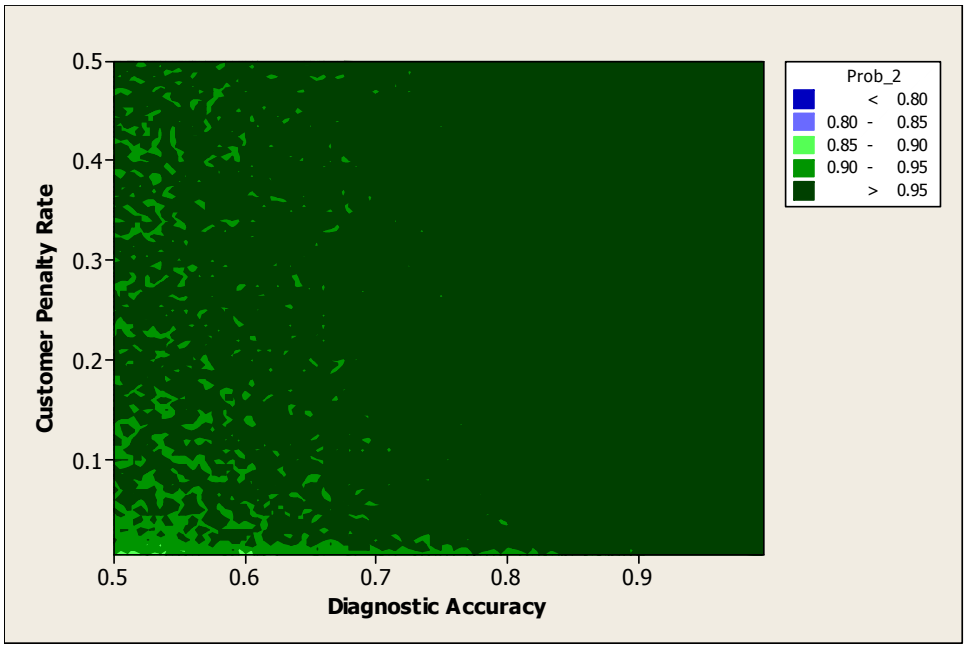


Figure 128: Contour plot of the probability of obtaining savings as a function of the customer penalty rate [cost units/customer/failure] and diagnostic accuracy for an urban region.

As Figure 128, Figure 127, Figure 128, and Table 57 all show, the probability of savings is quite high as, in general, at least 80% of the observed cases produce savings better than 90% of the time. In fact, in the urban case, fully 100% of the cases lead to better than 90% probabilities of savings.

Table 57: Distribution of cases amongst the probability ranges.

Probability Range	Rural Count [% of Cases]	Suburban Count [% of Cases]	Urban Count [% of Cases]
< 0.80	0.00	0.00	0.00
0.80 – 0.85	7.37	1.02	0.00
0.85 – 0.90	50.96	30.55	0.12
0.90 – 0.95	40.02	49.89	8.04
> 0.90	1.65	18.54	91.84

10.4.4 Value at risk and Probability of Loss

For the 1000 component examples, it is again useful to examine the value at risk and probability of loss for each of the region types. Table 58 shows the value at risk for each of the region types.

Table 58: Value at risk for each region type.

Region Type	Value at Risk [Cost Units]
Rural	1055
Suburban	1162
Urban	19837

As in the 100 component examples, the highest value at risk occurs again for the urban case. Combining this information with the probability of loss shown in Table 59 and it is clear that the trends observed in Section 10.3.4 also holds for these cases.

Table 59: Expected probability of loss for each region type.

Region Type	Expected Prob. of Loss
Rural	0.089
Suburban	0.087
Urban	0.026

10.5 REFINING THE RESULTS

One thing that must be considered is the fact that the computed probabilities are highly dependent on the diagnostic accuracy and failure rate ranges used in the simulation. These ranges can be further refined by weighting each input by its relative probability of occurring within the utility's system. For example, the diagnostic accuracies are not typically uniformly distributed over the interval [0.51, 0.99]. During the NEETRAC CDFI project, several diagnostic techniques have been found to produce overall accuracies of approximately 80% (±5%). One could weigh the probabilities plotted in the contour plots by the probability of a diagnostic accuracy and failure rate occurring. Figure 129 shows a sample PDF for diagnostic accuracies observed in underground cable system diagnostics.

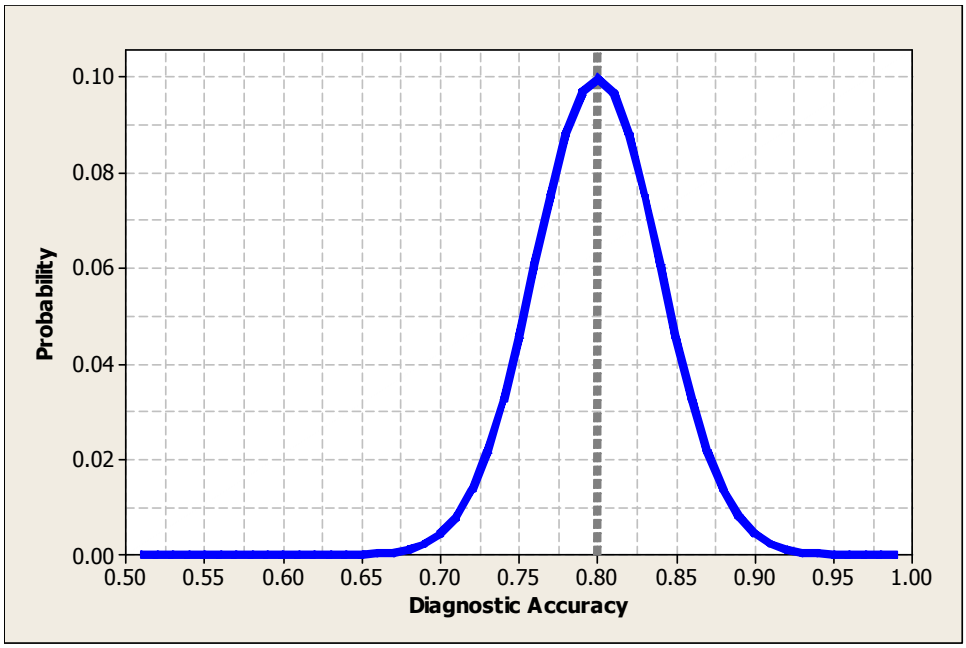


Figure 129: Sample PDF of diagnostic accuracies for an underground cable system diagnostic test.

Using Figure 129, it is then possible to adjust the weighting of for both the value at risk and probability of loss to account for what the likely accuracy of the diagnostic test will be. One can also do the same for the system failure rates as shown in Figure 130.

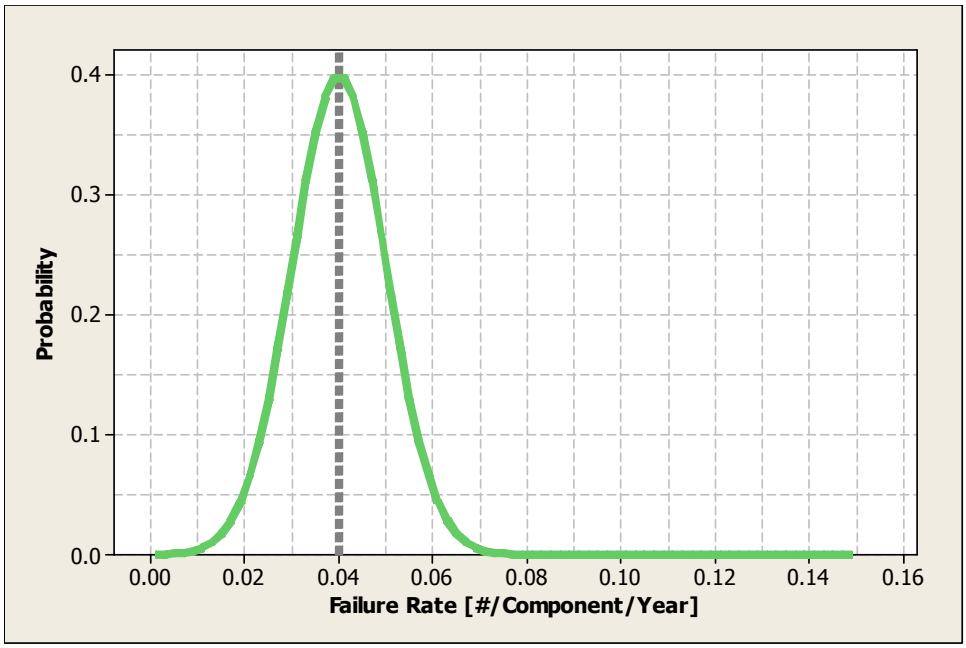


Figure 130: Sample PDF of local failure rates for different at-risk populations.

The PDFs shown in Figure 129 and Figure 130 are both normal distributions and so to combine them one can define a bivariate normal distribution as follows:

$$p(d,f,\rho)=\frac{1}{2\pi\sigma_d\sigma_f\sqrt{1-\rho^2}}\cdot\exp\left(Q(d,f,\rho)\right)$$

(10.1)

$$Q(d,f,\rho)=-\frac{1}{2\left(1-\rho^2\right)}\left(\frac{(d-\mu_d)^2}{\sigma_d^2}+\frac{\left(f-\mu_f\right)^2}{\sigma_f^2}-\frac{2\rho\left(f-\mu_f\right)(d-\mu_d)}{\sigma_d\sigma_f}\right),$$

where,

d = diagnostic accuracy,

f = failure rate,

ρ = correlation coefficient between the diagnostic accuracy and failure rate,

μ_d, σ_d = distribution mean and standard deviation for the distribution of diagnostic accuracies,

μ_d, σ_d = distribution mean and standard deviation for the distribution of diagnostic accuracies.

For the distributions in Figure 129 and Figure 130, the distribution parameters are shown in Table 60. Note that the correlation between diagnostic accuracy distributions and failure rate is assumed to be zero.

Table 60: Normal distribution parameters for diagnostic accuracies and failure rates.

Input	Parameter	Value
Diagnostic Accuracy	μ_d	0.80
	σ_d	0.04
Failure Rate	μ_f	0.04
	σ_f	0.01
Both	ρ	0

Figure 131 and Figure 132 shows the resulting surface and contour plots, respectively, for the bivariate normal distribution given by the parameters in Table 60.

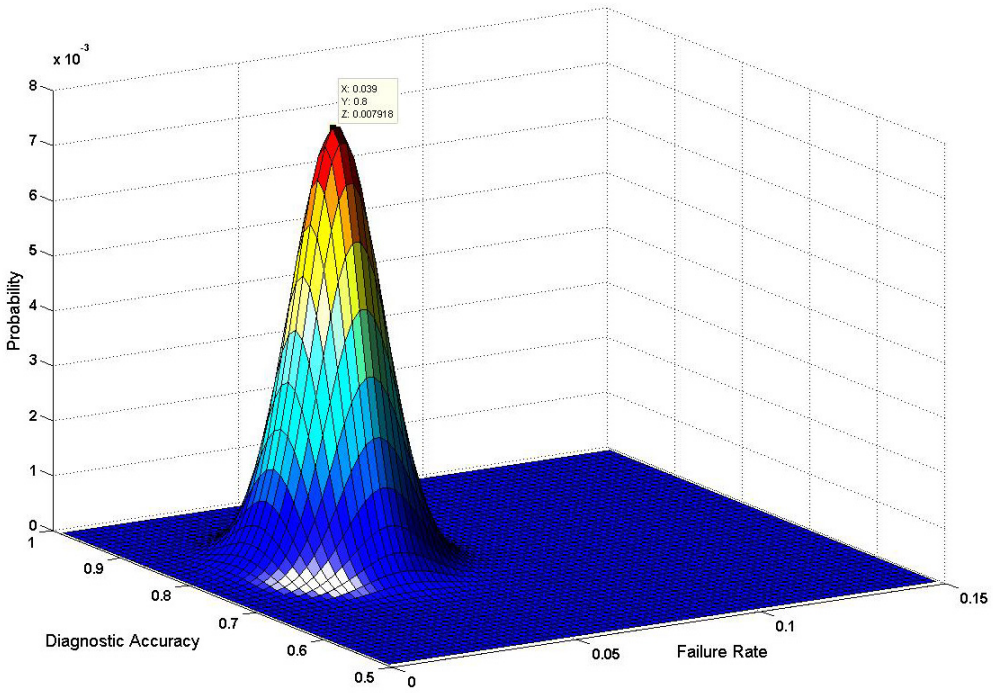


Figure 131: Surface plot showing the joint distribution of diagnostic accuracy and failure rate assuming the univariate distributions in Figure 129 and Figure 130.

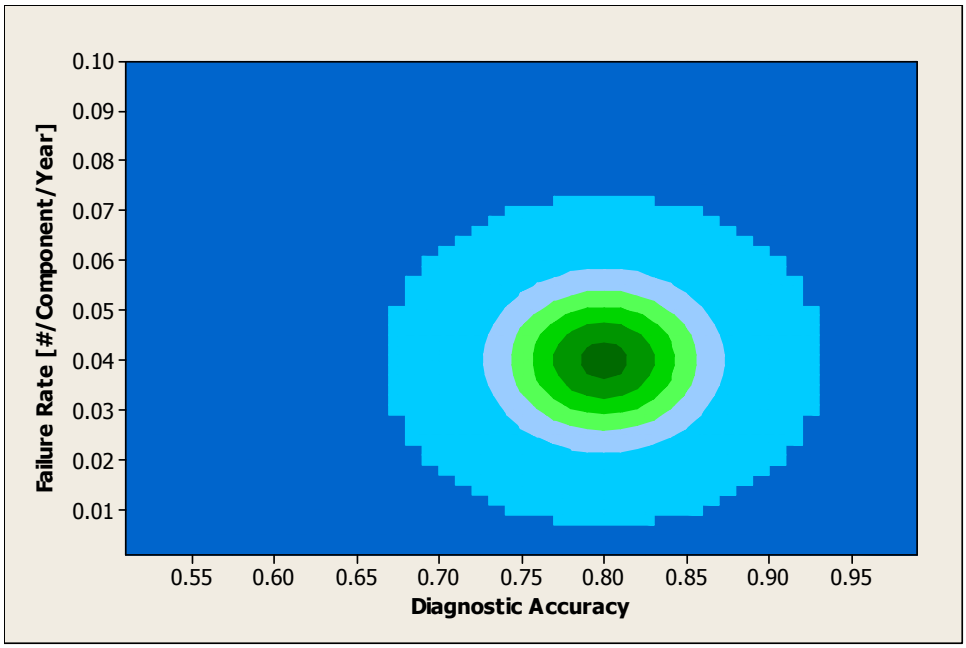


Figure 132: Contour version of Figure 131 with mean diagnostic accuracy of 0.8 and mean failure rate of 0.04.

Notice that the choice of the bivariate distribution has produced probabilities equal to zero for a large percentage of the failure rates and diagnostic accuracies initially used in the simulation. In fact, approximately 82.8% of the cases originally investigated would not occur if the bivariate distribution is assumed to represent the real system and chosen diagnostic. This is vital information to accurately assessing the value at risk and probability of loss that the program poses to the utility.

If the information in Figure 131 and Figure 132 is then applied to the weighting given to each value at risk and probability of loss, it is possible to obtain the likely values for these measures given the system and chosen diagnostic technique. This new weighting produces an expected value for each of the cases as shown in Table 61 and Table 62.

Table 61: Summary of the average probability of loss for each scenario considering the weights of Figure 131 and Figure 132.

	At Risk Population Size [Components]	
Region Type	**Prob. of Loss 100 [%]**	**Prob. of Loss 1000 [%]**
Rural	6.0	1.9
Suburban	2.5	1.9
Urban	2.0	1.9

Table 62: Summary of the value at risk for each scenario considering the weights of Figure 131 and Figure 132.

	At Risk Population Size [Components]	
Region Type	**Value at Risk 100 [Cost Units]**	**Value At Risk 1000 [Cost Units]**
Rural	14	1
Suburban	11	1
Urban	10	1

As Table 61 shows, loss probability levels for diagnostic programs performed on a system with the characteristics shown in Figure 131 and Figure 132 are very low. In fact, the maximum probability occurs for the 100 component rural region program and that probability is still only 6.0%. In other words, the expectation is that a savings will be generated in 94% of the possible cases. For most utilities, this should be acceptable. Unfortunately, the operating conditions in this example are not typical of the use of diagnostic programs by utilities. Alternative situations may arise such as the case described in Table 63.

Table 63: Normal distribution parameters for a second situation.

Input	Parameter	Value
Diagnostic Accuracy	μ	0.60
	σ	0.08
Failure Rate	μ	0.02
	σ	0.01
Both	ρ	0

The resulting distribution is shown in surface form in Figure 133 and in contour form in Figure 134. In this case, both the diagnostic accuracy and failure rate have been reduced as compared to the first example.

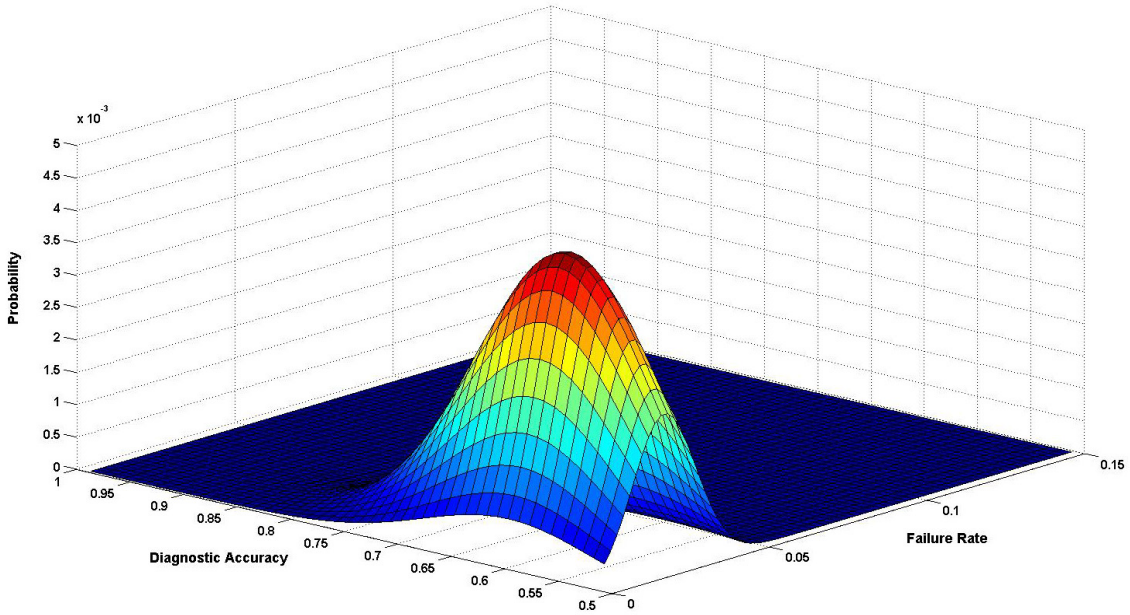


Figure 133: Surface plot of alternate diagnostic accuracy and failure rate scenario.

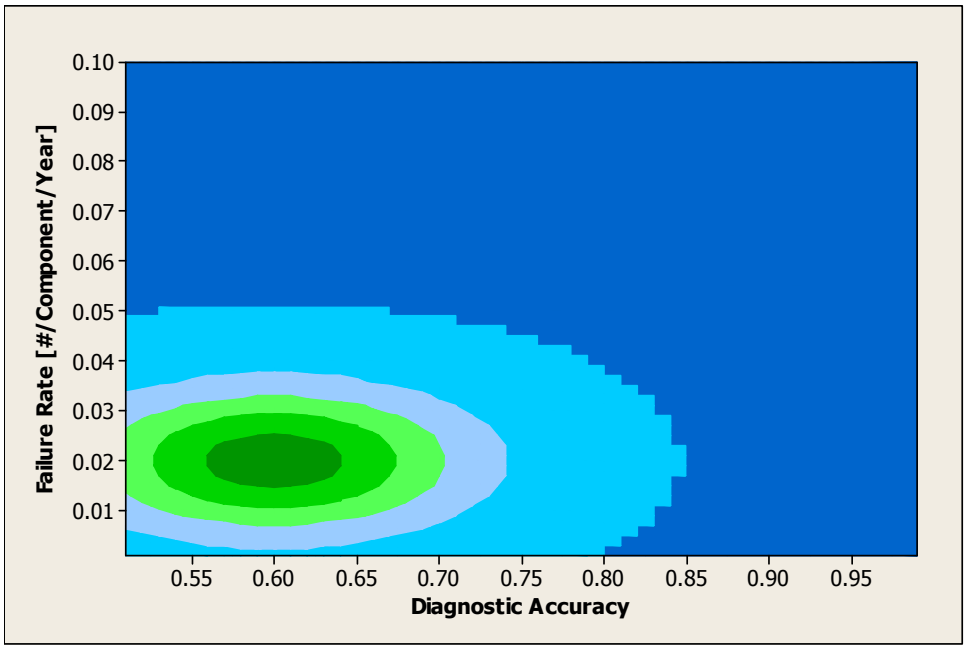


Figure 134: Contour version of Figure 133 with mean diagnostic accuracy of 0.6 and mean failure rate of 0.02.

Using Figure 133 and Figure 134 the expected value at risk for each of the region types is shown in Table 64.

Table 64: Summary of the probability of loss for each scenario considering the weights of Figure 133 and Figure 134.

	At Risk Population Size [Components]	
Region Type	**Prob. of Loss 100 [%]**	**Prob. of Loss 1000 [%]**
Rural	81.9	27.9
Suburban	42.8	20.5
Urban	18.6	6.8

Table 65: Summary of the value at risk for each scenario considering the weights of Figure 133 and Figure 134.

	At Risk Population Size [Components]	
Region Type	**Value at Risk 100 [Cost Units]**	**Value at Risk 1000 [Cost Units]**
Rural	635	1697
Suburban	843	2017
Urban	25064	29408

As compared to Table 61, Table 64 shows substantially higher probabilities of loss. By increasing the at-risk component population that probability can be reduced. However, an order of magnitude increase in population size only produces reductions of 66%, 52%, and 63%, for the rural, suburban, and urban regions, respectively. Only in the urban region case does the probability of loss drop below 10% and only for the 1000 component population.

The information presented in this section is critical to planning a diagnostic program as it can aid in the selection of a suitable at-risk population for a given diagnostic accuracy and risk. The goal in any such program is to reduce the risk as low as possible. However, as this analysis shows, the risk may still be substantial.

10.6 SUMMARY

This chapter demonstrates several key observations as to the behavior of the economics model developed in Chapter 9. These points are summarized below:

- At-risk populations with low failure rates involve greater risk than their high failure rate counterparts.
- Diagnostic accuracy can increase the likelihood of obtaining savings from low failure rate regions, but only marginally so.
- The average number of customers that may be impacted by a service outage has strong influence over the probability of savings. Higher numbers of customers tend to increase the probability.
- The expected value at risk can be more accurately computed if one considers the probability of each situation in the calculation.
- The customer penalty rate appears to be less critical than was first thought since it has little effect on the probabilities of generating a savings.

PART III SUMMARY

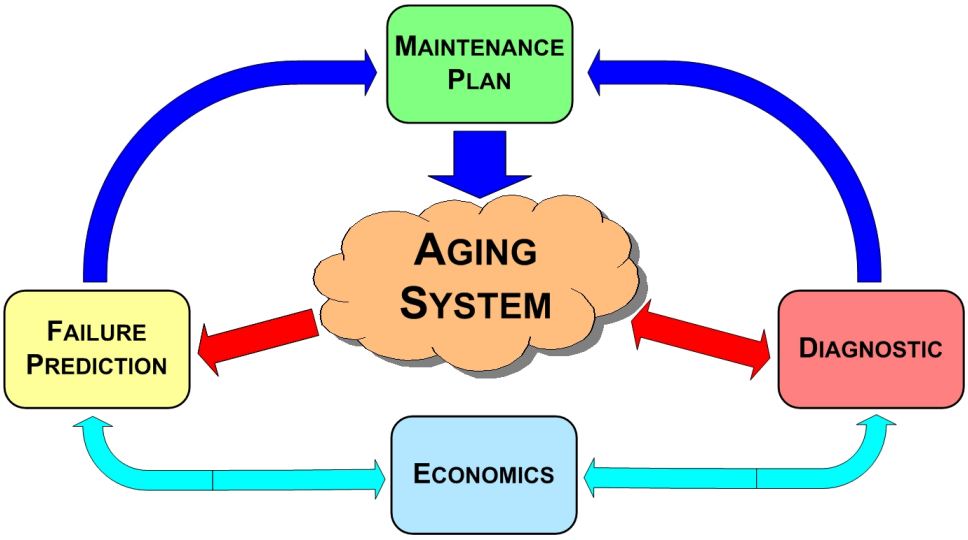
The failure prediction facet demonstrates that a diagnostic program may be decomposed into four stages: selection, action, generation, and evaluation. The total cost of the diagnostic program becomes the sum of the costs incurred during each of these stages. As Chapter 9 shows, the costs are finite and may be modeled using either deterministic values or random variables.

Once the total diagnostic program is computed, the calculation of savings requires only knowing what alternate program the diagnostic program should be compared to. The two cases that are most common in the power industry are complete replacement and "run to failure." In the first case, all components in the at-risk population are replaced. The diagnostic program, in this case, will produce savings as a result of the reduced number of corrective actions that are called for by the diagnostic program. In the case of "run to failure," this program calls for corrective action only once a component has failed in service. In this second case, the diagnostic program produces a savings through the improvements in reliability that result from removing "bad" components from service before they can fail. Other diagnostic programs may be compared in which case the resulting savings by one diagnostic program over the other would likely include contributions from both reduced spending on corrective actions and improved reliability.

The case studies in Chapter 10 demonstrate that diagnostic programs conducted in high failure rate areas have a high probability of producing saving as compared to a "run to failure" approach. In these high failure rate areas where the G/B ratio is worse than 85/15, the diagnostic accuracy has little impact on the probability of savings. Furthermore, at-risk populations located in urban regions also have a high probability of producing savings because of the high cost associated with each service failure. In these

cases, the concern is the high value at risk. Even though the probability is low that the program will produce a loss, the amount of loss that could be incurred is quite large as compared to the loss from a program conducted in a sparsely populated rural region. One must decided which risk is more crucial when planning the diagnostic program.

PART IV: CONCLUSIONS



CHAPTER 11: SUMMARY AND CONCLUSIONS

This chapter presents the summary and conclusions of the research work conducted over the course of the author's PhD program.

11.1 SUMMARY OF RESEARCH

The objective of this research has been to develop a statistical algorithm for controlling failure trends through the targeted maintenance of components that are at-risk for failure. Such work falls under the general heading of asset management. This work has required the development of three facets that are capable of addressing the key issues surrounding the deployment of diagnostic tests in maintenance programs. The following facets have been developed:

- Failure Prediction Facet – Two Weibull distribution-based prediction models that are able to predict service failures for a homogeneous population of components. Both algorithms can then be used to estimate the number of corrective actions that would be needed to address poor reliability within the population.
- Diagnostic Facet – Methods and techniques for assessing the accuracy and validity of diagnostic tests performed in the field on power system equipment.
- Economics Facet – Specification of the design process for general diagnostic programs and identification of the corresponding cost model for evaluating the economic benefit of diagnostic programs.

The key concepts from each facet are summarized in the following sections.

11.1.1 Failure Prediction Facet (Part I)

The failure prediction model encompasses two mathematical models, Model I and Model II, for predicting future failure rates based on the Weibull distribution. Each model

utilizes input data that must include information on the component population size and failure history. The main difference between the two is the need for additional data in the case of Model I in the form of components at the time of failure. Model II, on the other hand, utilizes failure data on the overall population of components. The data currently available in utilities is generally only suitable for Model II. However, it is expected that as utilities improve their data collection and storage that Model I will be the preferred choice.

Each model is able to predict the number of failures the utility might expect to see in the next several years. This process can then, in a sense, be reversed so that the models each estimate the number of corrective actions needed to alter the performance of the system by a desired amount. This allows the utility to establish both where the system is heading in terms of reliability and what level of action is needed to curtail an unacceptably high failure rate in the future.

These prediction models become the backbone of a stochastic simulation that is executed using Monte Carlo techniques. Using such an approach allows one to associate a probability with each prediction. The same applies to the calculation of the number of corrective actions needed to alter the population's predicted failure trend. This allows the utility to develop reliability improvement programs with sufficient resources to yield the desired improvement. Likewise, in systems that are still operating with high reliability, the failure prediction facet may be used to identify the time at which actions must be taken to prevent serious reductions in reliability performance.

Comparison of the two prediction models shows that Model I does provide greater efficiency in terms of allocating the corrective actions to different portions of the component population. Furthermore, the results of stochastic simulation show that the distributions of output variables for Model I are considerably narrower than those of Model II. In addition, the estimated number of corrective actions can be as much as 30%

less for Model I. This gives Model I a definite advantage. However, Model II is serves the same purpose quite effectively and until data become available with the level of detail required by Model I, Model II is a reasonable alternative.

The failure prediction facet can address four key questions in the asset management chain:

- What portion of the system has too large a failure rate?

  The failure prediction facet can provide area by area predictions of failure rates. In addition, it is able to identify what component vintages are contributing most to the current and future failure rates provided the data to do so are available.

- What amount of resources will be needed to reduce the failure rate?

  An estimate of the required number of corrective actions that are needed to alter the future failure rate may be estimated using the failure prediction facet. This provides the utility with a clear idea of the size of diagnostic program that will be needed to address the high failure rates.

- When does the utility need to start targeting the high failure rate areas?

  In systems where the failure rate has yet to reach high enough levels to warrant immediate action by the utility, the predictions can be used to determine how long the utility has before the failure rates do get high enough.

The diagnostic and economics facets each address other issues in the asset management chain.

11.1.2 Diagnostic Facet (Part II)

The diagnostic facet provides the means of assessing the usefulness of diagnostic testing techniques employed in the field. The main interest is determining the overall diagnostic accuracy that a particular diagnostic technique can deliver (Chapter 6). This is a deceptively complex process that has required the use and development of several

analytical techniques. The plethora of diagnostic techniques and their implications make the comparison process difficult without these specialized analysis tools. These analysis techniques include performance ranking, diagnostic outcome mapping, Weibull analysis, survivor analysis, and classifiers (Chapter 7). Each of these techniques is a necessary element in the interpretation of diagnostic measurements performed on components presently operating in the field.

This facet allows one to answer the following questions:

- What is the risk of failure during the diagnostic test?

  Failures on test occur occasionally with diagnostic tests that employ test voltages that are higher than the operating voltage of the component. These failures on tests can be evaluated directly using techniques such as Weibull analysis and survivor analysis.

- What do diagnostic provider recommendations mean?

  Correlating diagnostic provider recommendation with service performance is best handled by the performance ranking and Weibull analysis techniques. These techniques provide information of the future service performance of components diagnosed by the diagnostic provider. This assumes that previous test programs have been performed in which components are diagnosed and then left in service and monitored for several years after the test.

- What is the probability of failure for components diagnosed in each condition?

  Weibull analysis can be applied to the data supplied by diagnostic providers to provide expectations as to the numbers of components that will fail by a certain time after the test. This will likely differ between component groups diagnosed with different conditions.

- What is the accuracy of the diagnostic test?

The most critical factor in diagnostic testing is the accuracy of the test. This is best assessed using a probabilistic approach that can adjust for different time periods since testing. If this data is not available, then the assumption must be made that all the components diagnosed as bad will have failed by the time the analysis is performed. This is most likely not realistic but it useful if no additional information is available.

These probabilities are vital components to the calculation of economic benefit as covered by the economics facet.

11.1.3 Economics Facet (Part III)

The economics facet differs from the other two facets because its focus is on the cost elements of the diagnostic program applied to an at-risk population of components. Specifically, this facet seeks to determine under what conditions different diagnostic programs can be employed to achieve the best possible, if any, economic benefit (savings). These savings are derived from reduced spending on corrective actions, improvements in reliability, or a combination of both. In order to compute these costs, the four stages of a diagnostic program have been defined and include selection, action, generation, and evaluation. Using these four stages, one can define a cost component that represents each stage and then sum them together to obtain the total cost.

Unfortunately, many of the cost function inputs are not known precisely so they must be treated as random variables. This leads to the use of stochastic simulation and, like the failure prediction facet, the cost function can be modeled using Monte Carlo techniques. This provides the probability that a diagnostic program will generate a savings as compared to an alternative program. Common alternatives include the complete replacement and “run to failure” approaches. The former requires the utility to replace all the components in the at-risk population while the latter implies that utility only performs corrective actions on those components that fail in service.

These two programs represent the extremes for action and also the extreme points in terms of savings. In the case of the complete replacement approach, the savings come directly from the diagnostic program's reduced spending on corrective actions. On the other hand, the savings as compared to the "run to failure" approach come exclusively from the improvement in reliability. Other programs will lie somewhere between these cases with some savings being produced from both the corrective actions and reliability improvement.

This facet relies on information from both the failure prediction and diagnostic facets. In the first case, the failure prediction facet aids the economics facet in identifying the at-risk population that will be used in the diagnostic program. This identification process is based, in theory, on the local failure rates as estimated from historical failure records. On the other hand, the diagnostic facet provides information on the accuracy of the different diagnostic tests and the effect these accuracies have on the results of the diagnostic testing. As shown in Chapter 10, the inputs from both of these facets are vital components to the results of the cost function.

Together, the three facets represent the key elements needed to generate a maintenance plan for addressing the need for utilities to manage their aging infrastructure.

11.2 CONCLUSIONS

The work presented on each of the three facets has revealed a number of key conclusions and observations as described below:

- **Failure Prediction Facet (Part I)**
    - Simulation results (both point and stochastic) show that Model I generates more efficient allocation of corrective actions than Model II (Chapter 3 and Chapter 5).

  - Both failure prediction models are able to model datasets in which the characteristic life may or may not be observed within the data itself (Chapter 3 and Chapter 5).
  - Results from underground cable field data shows that the choice of the Weibull distribution is reasonable.

- **Diagnostic Facet (Part II)**
  - Two types of diagnostic accuracies are needed: (1) overall diagnostic accuracy and (2) condition-specific diagnostic accuracy.
  - Diagnostic accuracies should be computed using a probabilistic approach that considers the time since test.
  - Diagnostic testing results must be validated using analysis tools such as performance ranking, diagnostic outcome mapping, Weibull analysis, and survivor analysis.
  - Each diagnostic test must be evaluated in order to ascertain its potential to provide reliable condition assessments.
- **Economics Facet (Part III)**
  - Diagnostic programs may be modeled using a four stage process: selection, action, generation, and evaluation (SAGE).
  - The cost components of diagnostic programs may be computed. However, several are best represented by random variables.
  - Diagnostic programs may produce economic saving from improved reliability and/or reduced spending on corrective actions.
  - The maximization of the economic benefits represents a non-linear stochastic optimization problem.
  - The probability of savings is strongly dependent on the at-risk population failure rate.

- Results of simulation studies indicate that there is a high probability of achieving a net savings in regions with high failure rates and/or high numbers of customers. The at-risk population in general should be composed of no more than 90% "good" components.
- High diagnostic accuracy is needed primarily in at-risk populations that are composed of more than 90% "good" components.
- Larger size at-risk populations generally lead to higher probabilities of savings (> 90%). However, larger populations also increase the value at risk.
- Diagnostic programs must balance spending on corrective actions against the potential improvements in reliability.

Based on this work, it is clear that employing diagnostic tests in the management of an aging infrastructure has the potential to provide economic savings as compared to the maintenance approaches currently employed by utilities. In order to take advantage of what these technologies have to offer, utilities must work towards assembling the data that represents their systems. This must be combined with the work currently undertaken to assess the accuracy of diagnostic techniques, such as the NEETRAC CDFI project in the case of underground cable systems.

Diagnostic programs require careful planning and assessment of their potential impact both on the performance of the power system. To produce an economic savings, these programs must be implemented in the right place, at the right time, and using the proper diagnostic and corrective actions. Utilizing the facets developed as part of this research will aid utilities in this endeavor.

11.3 CONTRIBUTIONS

The completed research has yielded the following contributions:

- Stochastic Failure Prediction Algorithm – A statistical approach to predicting future failure trends based on data either as limited as annual installs, replacements, and cumulative failures (i.e. partial information), or data that includes the ages of failed components (complete information). These predictions may then be altered by adjusting the component population through targeted maintenance actions that restore the worst performing components to new working order.
- Formulation of Monte Carlo from Population Distribution – Developed methodology for generating randomized datasets for Monte Carlo simulation based on the bulk population distribution. This PDF is generated either from data as limited as annual installation, replacements, and cumulative failures, or from data that includes the ages of failed components. For each year of data, a separate distribution is produced based on the population distribution but modified to account for the original distribution's expected number of failures for that year.
- Methodologies for Evaluating Diagnostic Testing Procedures – A series of techniques (performance ranking and diagnostic outcome mapping) that can cope with the wide range of diagnostic data possibilities and limited availability of data. These techniques ultimately allow for calculation of the accuracies of different diagnostic testing procedures for input into the economics model.
- Diagnostic Program Cost Model – An economic model that takes into consideration the variety of maintenance actions that may be performed in connection with different diagnostic testing procedures. More importantly, though, this model performs these calculations even when the diagnostic accuracy is imperfect (less than 100% accurate). The model is able to account for the consequences that result from the imperfections in the diagnostic. The cost components themselves become linear or nonlinear combinations of random

variables from which the distribution of total savings may be computed given known distributions for these components.

- Definition of Diagnostic Directed Maintenance Plan – A four stage framework for implementation and evaluation of a diagnostic program. These stages include selection, action, generation, and evaluation. This framework guides a utility through the process of choosing components to test, deciding what maintenance actions to perform given results of the testing, selection of the diagnostic technique, and evaluation of the program performance. A diagnostic program is a living process that can be continuously improved through this framework.
- Revisions to IEEE Standards – Through the collection and analysis of diagnostic and service data valuable knowledge has been gained that is in contrast to existing IEEE standards. As these standards have entered into their periods for revision this information is being considered as part of those revisions.

11.4 FUTURE WORK

The future work in this field of research may be approached from the perspectives of basic research and applications as discussed in the following sections.

11.4.1 Basic Research

The primary opportunity for basic research in this field is the development of the solution to the stochastic optimization of the economic cost function. As described in Section 9.4.3, required stochastic optimization process is non-linear in nature. Techniques should be investigated that would allow one to obtain a solution to this problem. The issues to consider include:

- Choice of decision variables.
- Definition of optimization objective (i.e., expected benefit, expected loss, expected value at risk, etc).

- Convergence and existence of solution.

These issues taken together make the solution difficult to obtain. Further work in this area would aid utilities in identifying the best way to proceed with their diagnostic programs.

11.4.2 Applications

The following application-based research would also aid utilities in implementing their diagnostic programs:

- Compilation of actual diagnostic program case studies.
- Development of a software tool that would allow utilities to conduct their own studies of the economics of diagnostic programs.
- Construction of a database of diagnostic accuracies that would allow for fast identification of diagnostic techniques with suitable overall accuracies.
- Planning and implementation of a real diagnostic program using the diagnostic program process described in Chapter 8.
- Development of a standardized database for maintaining component information including installation, failure history, diagnostic testing results, utility expenditures on that component, annual SAIDI/SAIFI contributions.

11.5 INDUSTRY CONTRIBUTIONS

This section reviews the industry contributions (publications) that have resulted from this research.

11.5.1 Patents

- Perkel, J.; Hampton, N; Begovic, M., "Methods to Determine the Accuracy / Validity of Diagnostic Techniques through Comparison with Performance Data," US Provisional Patent 60/888,162. Submitted February 5, 2007.

11.5.2 Journal Papers

- Hernández-Mejía, J.C., Perkel, J., Harley, R., Begovic, M., Hampton, N., and Hartlein, R., "Determining Routes for the Analysis of Partial Discharge Signals Derived from the Field," Accepted for publication in the *IEEE Transactions on Dielectrics and Electrical Insulation*.
- Begovic, M; Perkel, J; Hartlein, R, "Equipment Failure Forecasting Based on Past Failure Performance and Development of Replacement Strategies," *KIEEME Transactions on Electrical and Electronic Materials*, Vol. 7, No. 5, pp. 217-223. Published October 2006.

11.5.3 Conference Papers Published

- Begovic, M.; Hampton, N.; Hartlein, R.; Perkel, J., Validation of the accuracy of practical diagnostic tests for power equipment," *Cigre 2008 Session D1*, Published August 2008, Paris, France.
- Hartlein, R.A.; Hampton, R.N.; Perkel, J., "Some Considerations on the Selection of Optimum Location, Timing, and Technique, for Diagnostic Tests," *2008 PES General Meeting*, Published July 2008, Pittsburgh, PA.
- Perkel, J.; Hampton, R.N.; Begovic, M.; Hartlein, R., "Validating 'Diagnostic Tests' for Cables in Service," *7th International Conference on Insulated Power Cables 2007 (JICABLE'07)*, Published June 2007, pp. 449-454, Paris, France.
- Perkel, J.; Begovic, M.; Hartlein, R., "Generating Cable Replacement Strategies Using Monte Carlo Simulation**,"** *Power Systems Conference and Exposition, 2006. PSCE '06. 2006 IEEE PES*, pp. 2188-2193, Published Oct. 29-Nov. 1, 2006, Atlanta, GA.
- Djuric, P.M.; Begovic, M.M.; Perkel, J., "Prediction of power equipment failures based on chronological failure records,**"** *Circuits and Systems, 2006. ISCAS 2006.*

Proceedings. 2006 IEEE International Symposium on, pp. 1207-1210 Published May 2006, Kos, Greece.

- Begovic, M.; Djuric, P.; Perkel, J.; Vidakovic, B.; Novosel, D., "New Probabilistic Method for Estimation of Equipment Failures and Development of Replacement Strategies," *System Sciences, 2006. HICSS '06. Proceedings of the 39th Annual Hawaii International Conference on*, Vol.10, pp. 246a- 246a, Published Jan. 2006, Kauai, HI.

11.5.4 Presentations

- Perkel, J., "Accuracies and Techno-Economics," EPRI Industry Summit: Advanced Diagnostics for Distribution Cables. Presented June 25, 2008, Chicago, IL.
- Hampton, R.N.; Perkel, J., "Analysis of VLF Field Data from Service," NEETRAC Cable Diagnostic Focused Initiative Update Meeting. Presented February 12, 2008, Atlanta, GA.
- Hampton, R.N.; Perkel, J., "Analysis of PD Field Data," NEETRAC Cable Diagnostic Focused Initiative Update Meeting. Presented February 11, 2008, Atlanta, GA.
- Hampton, N.; Perkel, J., "Benefit/Value of Diagnostic Testing," NEETRAC Cable Diagnostic Focused Initiative Update Meeting. Presented February 11, 2008, Atlanta, GA.
- Perkel, J; Begovic, M.; Hampton, N.; Hartlein, R., "Employing Diagnostics to Manage Underground Assets," *IEEE PES General Meeting*. Poster presented June 24-28, 2007, Tampa, FL.

- Hampton, N.; Perkel, J., "VLF Diagnostics," *IEEE PES Insulated Conductors Committee Meeting (ICC), C18D Discussion Group, Spring Meeting*. Presented May 8, 2007, Orlando, FL.
- Perkel, J., "Validating the Accuracy of Diagnostics," NEETRAC Cable Diagnostic Focused Initiative Update Meeting. Presented February 7, 2007, Atlanta, GA.
- Hampton, N.; Perkel, J., "Real Test Program: Evaluation and Suggested Improvements," NEETRAC Cable Diagnostic Focused Initiative Update Meeting. Presented February 8, 2007, Atlanta, GA.
- Begovic, M.; Perkel, J., "Experience of Withstand Testing," NEETRAC Cable Diagnostic Focused Initiative Update Meeting. Presented February 8, 2007, Atlanta, GA.
- Begovic, M.; Hampton, R.N.; Hartlein, R.; Perkel, J., "Validating Cable 'Diagnostic Tests'," Presented at *IEEE Insulated Conductors Committee (ICC) Meeting*, *Subcommittee C, Fall Meeting*. Presented October 2006, St. Petersburg, FL.
- Andrews, T.; Begovic, M.; Hampton, R.N.; Hartlein, R.; Harley, R.; Hernandez, J.C.; Parker, T; Perkel, J.; "Reflections on 'First Experiences' with IEEE 400," Presented at *IEEE Insulated Conductors Committee (ICC) Meeting, WG C-18, Fall Meeting*. Presented October 2006, St. Petersburg, FL.
- Begovic, M.; Hampton, N.; Perkel, J., "Activity 3: Analysis of Existing Data," NEETRAC Cable Diagnostic Focused Initiative Update Meeting. Presented May 4, 2006, Atlanta, GA.
- Perkel, J; Begovic, M., "Equipment Asset Management: A Failure Performance Based Approach," *IEEE PES General Meeting*. Presented June 2005, San Francisco, CA.

11.5.5 Software

- Begovic, M.; Hartlein, R.; Patel, N.; Perkel, J., "Cable Asset Management Software," NEETRAC Project 05-1991, Distributed May 2006.

APPENDIX A

FAILURE DATA

As described in Chapter 2, the two available failure prediction models each require differing levels of detail in their input data. Unfortunately, no dataset is readily available with the detail needed for Model I. In lieu of this, this chapter will describe the procedure for generating two datasets that will be compatible with both failure prediction models.

A.1 SYNTHESIZING DATASETS

Given the unavailability of field data with sufficient detail to utilize Model I, two synthesized datasets have been generated to demonstrate the differences between Model I and Model II. The data used in Model II is a subset of the data used in Model I. Therefore, the process is to generate the dataset for Model I and then extract a subset of that information for use in Model II. Table 66describes the inputs for each model.

Table 66: Input data for each model.

Model	Input	Description
--	N	Total number of new component installations performed each year.
I	F	Total number of component failures occurring in year j from components originally installed in year i ($j \geq i$).
	X	Total number of components installed in year i still in service in year j ($j \geq i$).
II	R_T	Total number of components replaced each year.
	F_T	Total number of component failures each year.

Note that the new installation input is the same for both models.

The datasets themselves are then generated from a specified Weibull distribution according to the following procedure:

1. Select the length of the dataset in years (n).
2. Select the number of new component installations (X_{new}) for each year represented in the population matrix, X, as the main diagonal as,

$$X = \begin{bmatrix} X_{1,1} & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & X_{2,2} & 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & X_{3,3} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & X_{4,4} & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & X_{5,5} & 0 & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & X_{n-1,n-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_{n,n} \end{bmatrix}. \quad \textbf{(A.1)}$$

3. Select the Weibull scale (α) and shape (β) parameters.
4. For each vintage *i,* randomly generate $X_{new(i)}$ times to failure (F_{ij}) using a Weibull distribution with parameters α and β.
5. Round each F_{ij} to the next integer value. This corresponds to the year during which the failure occurred. For example, for a failure time of 1.4 years, the year of failure is computed as two since the failure occurred during the second year.
6. For each vintage *i*, determine the total number of failures during each year for all years less than *n-i* as,

$$f = \begin{bmatrix} 0 & f_{1,2} & f_{1,3} & f_{1,4} & \cdots & \cdots & \cdots & f_{1,n} \\ 0 & 0 & f_{2,3} & f_{2,4} & f_{2,5} & \cdots & \cdots & f_{2,n} \\ 0 & 0 & 0 & f_{3,4} & f_{3,5} & \cdots & \cdots & f_{3,n} \\ 0 & 0 & 0 & 0 & f_{4,5} & \cdots & \cdots & f_{4,n} \\ 0 & 0 & 0 & 0 & 0 & f_{5,6} & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & f_{n-2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & f_{n-1,n} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad \textbf{(A.2)}$$

7. Adjust the population by subtracting failed components from the appropriate vintage and then adding them as new installations for that year. This process is shown in (A.3) and completes the synthesis of data for Model I,

$$X_{i,j} = X_{i,j-1} - f_{i,j}, \quad j > i,$$

$$X_{j,j} = X_{new}(j) + \sum_{i=1}^{j-1} f_{i,j}. \tag{A.3}$$

8. Extract from f the data needed for Model II as,

$$F_T(j) = \sum_{i=1}^{j-1} f_{i,j},$$

$$R_T(j) = F_T(j). \tag{A.4}$$

Each synthesized dataset will include 20 years worth of data with each input representing a matrix or vector of the sizes shown in Table 67.

Table 67: Matrix dimensions for each input variable.

Model	Input	Matrix Dimensions	
		Rows [#]	Columns [#]
--	X_{new}	20	1
I	F	20	20
	X	20	20
II	F_T	20	1
	R_T	20	1

Two datasets will be synthesized using different Weibull parameters that correspond to different characteristic lifetimes. Dataset 1 will include components with characteristic lifetimes of 30 years while Dataset II will include components with characteristic lifetimes of only 20 years. These datasets are each described in detail in Sections A.2 and A.3.

A.2 DATASET 1 – LONG COMPONENT CHARACTERISTIC LIFETIME

Dataset 1 is constructed using the installation input, *N*, as shown in Table 68 and the Weibull parameters (α and β) as shown in Table 69. Note that the annual installations would follow a pattern of introduction (phase in), routine installation, transition to different component design (phase out), and finally maintenance of existing components.

Table 68: Input vector *N* of new installations.

Year	New Installations [Components]	Installation Mode
1	100	Phase in
2	290	
3	250	
4	450	Routine Installation
5	500	
6	630	
7	500	
8	390	Phase Out
9	240	
10	150	
11	0	Maintain Operation
12	0	
13	0	
14	0	
15	0	
16	0	
17	0	
18	0	
19	0	
20	0	

Table 69: Weibull parameters used to generate Dataset 1.

Parameter	Value
α	30
β	1.5

The resulting datasets for each model are shown in Sections A.2.1 and A.2.2.

A.2.1 Raw Data for Model I

Table 70 and Table 71 show the resulting matrices for F (annual failures) and X (population), respectively.

Table 70: Annual failure data (*F*) for each vintage of Dataset 1.

	Year																			
Vintage	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	0	1	1	0	4	3	0	2	2	5	2	4	2	6	3	1	3	4	4	2
2	0	0	1	4	5	5	3	8	10	8	5	3	4	9	7	9	9	6	8	6
3	0	0	0	0	3	3	3	4	3	6	4	9	7	6	5	7	7	9	5	5
4	0	0	0	0	3	6	9	6	10	4	12	8	7	7	13	13	4	14	13	12
5	0	0	0	0	0	5	7	3	11	9	9	11	11	11	9	12	11	12	10	9
6	0	0	0	0	0	0	4	8	9	8	11	11	18	13	15	10	13	14	15	16
7	0	0	0	0	0	0	0	5	11	5	11	12	13	13	14	15	10	7	13	16
8	0	0	0	0	0	0	0	0	4	5	5	14	7	8	16	9	13	9	20	12
9	0	0	0	0	0	0	0	0	0	3	3	4	5	4	4	4	5	6	4	5
10	0	0	0	0	0	0	0	0	0	0	5	1	2	3	4	1	0	3	3	7
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	3	2	4
12	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	2
13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	2	0	1	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	3
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	1
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 71: Population remaining in service each year (*X*) of Dataset 1.

	Year																			
Vintage	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	100	99	98	98	94	91	91	89	87	82	80	76	74	68	65	64	61	57	53	51
2	0	291	290	286	281	276	273	265	255	247	242	239	235	226	219	210	201	195	187	181
3	0	0	252	252	249	246	243	239	236	230	226	217	210	204	199	192	185	176	171	166
4	0	0	0	454	451	445	436	430	420	416	404	396	389	382	369	356	352	338	325	313
5	0	0	0	0	515	510	503	500	489	480	471	460	449	438	429	417	406	394	384	375
6	0	0	0	0	0	652	648	640	631	623	612	601	583	570	555	545	532	518	503	487
7	0	0	0	0	0	0	526	521	510	505	494	482	469	456	442	427	417	410	397	381
8	0	0	0	0	0	0	0	426	422	417	412	398	391	383	367	358	345	336	316	304
9	0	0	0	0	0	0	0	0	300	297	294	290	285	281	277	273	268	262	258	253
10	0	0	0	0	0	0	0	0	0	203	198	197	195	192	188	187	187	184	181	174
11	0	0	0	0	0	0	0	0	0	0	67	67	67	67	64	63	63	60	58	54
12	0	0	0	0	0	0	0	0	0	0	0	77	77	76	76	75	74	74	74	72
13	0	0	0	0	0	0	0	0	0	0	0	0	76	75	74	73	71	71	70	69
14	0	0	0	0	0	0	0	0	0	0	0	0	0	82	82	81	81	81	79	76
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	94	94	94	94	93	91
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	85	84	84	80	79
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	79	79	79	76
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87	87	86
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	105	105
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	107

A.2.2 Reduced Data for Model II

The extraction from the Model I dataset yields the inputs R_T and F_T shown in Table 72 for use in Model II.

Table 72: Complete Model II data extracted from Dataset 1 for Model I.

Year	X_{new} [Components]	R_T [Components]	F_T [Components]
1	100	0	0
2	290	1	1
3	250	2	2
4	450	4	4
5	500	15	15
6	630	22	22
7	500	26	26
8	390	36	36
9	240	60	60
10	150	53	53
11	0	67	67
12	0	77	77
13	0	76	76
14	0	82	82
15	0	94	94
16	0	85	85
17	0	79	79
18	0	87	87
19	0	105	105
20	0	107	107

A.2.3 Dataset Properties

The Weibull parameters used to generate the dataset imply that the component population displays two key characteristics:

- The components are in the aging region of the Weibull bathtub curve (see Section 7.3 for additional details) since the Weibull shape parameter is greater than one.
- The characteristic lifetime defined as the age at which 63.2% of the components will fail is 30 years.

The first point implies that over time the failure rate for the population will increase as the components age. This is the definition of "aging." On the other hand, the second characteristic implies that the components have a longer life than could have been observed during the 20 years of data. This is quite common in power system equipment as records have only recently begun to be maintained.

The resulting annual failures are shown in Figure 135.

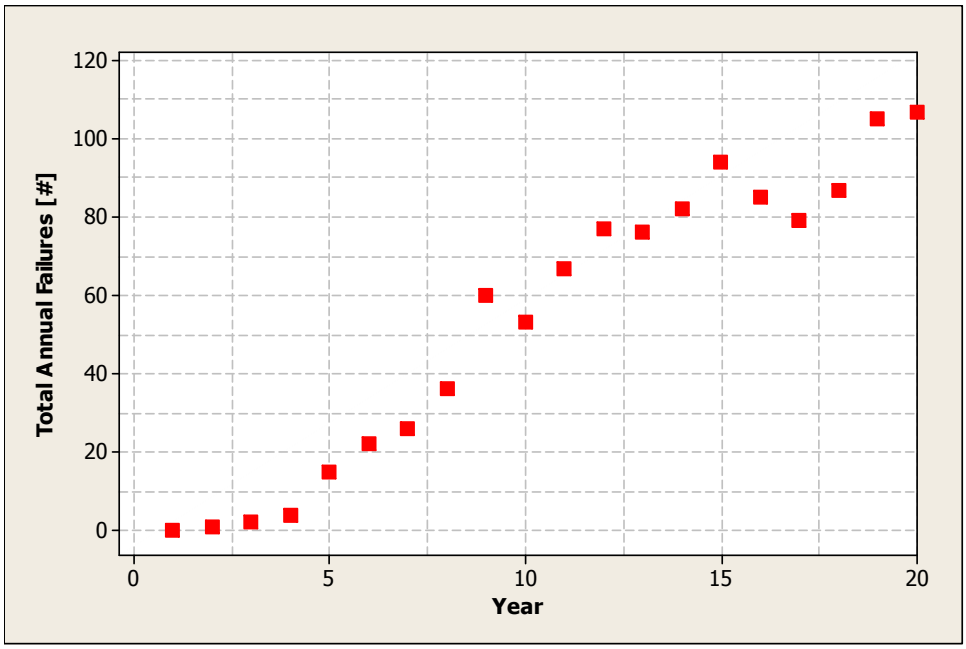


Figure 135: Annual failures for Dataset 1.

The combination of total annual failures and installed population lead to the annual failure rates shown in Table 73. Note that after new installations cease the dataset shows the total population of in service components remains constant. This is consistent with the maintenance phase shown in Table 68.

Table 73: Annual failure rates for Dataset 1.

Year	New Installations [Components]	In Service Population [Components]	Annual Failures [#]	Annual Failure Rate [#/100 Comp./Year]
1	100	100	0	0.00
2	290	390	1	0.26
3	250	640	2	0.31
4	450	1090	4	0.37
5	500	1590	15	0.94
6	630	2220	22	0.99
7	500	2720	26	0.96
8	390	3110	36	1.16
9	240	3350	60	1.79
10	150	3500	53	1.51
11	0	3500	67	1.91
12	0	3500	77	2.20
13	0	3500	76	2.17
14	0	3500	82	2.34
15	0	3500	94	2.69
16	0	3500	85	2.43
17	0	3500	79	2.26
18	0	3500	87	2.49
19	0	3500	105	3.00
20	0	3500	107	3.06

Table 73 also provides further evidence that the population is aging since the failure rate displays a generally increasing trend over time.

A.3 DATASET 2 – SHORT COMPONENT CHARACTERISTIC LIFETIME

The construction of Dataset 2 is similar to that of Dataset 1 as it uses the same new installation input, *N*, as shown in Table 68. The difference between the two datasets is the choice of Weibull parameters (α and β) as shown in Table 74.

Table 74: Weibull parameters used to generate Dataset 2 as compared to the parameters for Dataset 1.

Dataset	Parameter	Value
1	α	30
	β	1.5
2	α	20
	β	2

Note that Dataset 2 has a short characteristic lifetime of 20 years as compared to the 30 years of Dataset 1 and a higher failure rate. This implies that 20 year duration of Dataset 2 is long enough to observe significant failures from the installed population. Indeed, it is possible that some vintages may be completely removed during that period.

A.3.1 Raw Data for Model I

Table 75 and Table 76 show the resulting matrices for *F* (annual failures) and *X* (population), respectively.

Table 75: Annual failure data (*F*) for each vintage of Dataset 2.

	Year																			
Vintage	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	0	0	2	2	0	3	3	6	1	2	6	0	7	5	8	3	6	4	6	3
2	0	0	2	1	4	4	9	3	10	8	14	14	14	9	13	12	14	14	4	11
3	0	0	0	1	1	6	1	4	4	12	11	9	9	9	8	10	10	11	11	13
4	0	0	0	0	0	3	6	8	13	11	10	14	16	15	30	13	22	14	18	24
5	0	0	0	0	0	0	3	8	8	15	12	12	20	19	15	23	12	28	17	19
6	0	0	0	0	0	0	2	3	7	12	15	23	18	13	20	27	24	29	31	19
7	0	0	0	0	0	0	0	0	4	4	8	15	15	6	19	21	18	22	16	21
8	0	0	0	0	0	0	0	0	1	2	3	7	10	11	15	12	11	11	22	16
9	0	0	0	0	0	0	0	0	0	0	3	3	10	7	7	9	12	11	7	12
10	0	0	0	0	0	0	0	0	0	0	0	2	3	3	7	4	3	9	10	9
11	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	3	0	2	2	1
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	3	3	4
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	4	3	6
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	5	1
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	2	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	2	4
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 76: Population remaining in service each year (*X*) of Dataset 2.

	Year																			
Vintage	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	100	100	98	96	96	93	90	84	83	81	75	75	68	63	55	52	46	42	36	33
2	0	290	288	287	283	279	270	267	257	249	235	221	207	198	185	173	159	145	141	130
3	0	0	254	253	252	246	245	241	237	225	214	205	196	187	179	169	159	148	137	124
4	0	0	0	454	454	451	445	437	424	413	403	389	373	358	328	315	293	279	261	237
5	0	0	0	0	505	505	502	494	486	471	459	447	427	408	393	370	358	330	313	294
6	0	0	0	0	0	646	644	641	634	622	607	584	566	553	533	506	482	453	422	403
7	0	0	0	0	0	0	524	524	520	516	508	493	478	472	453	432	414	392	376	355
8	0	0	0	0	0	0	0	422	421	419	416	409	399	388	373	361	350	339	317	301
9	0	0	0	0	0	0	0	0	288	288	285	282	272	265	258	249	237	226	219	207
10	0	0	0	0	0	0	0	0	0	216	216	214	211	208	201	197	194	185	175	166
11	0	0	0	0	0	0	0	0	0	0	82	82	81	80	80	77	77	75	73	72
12	0	0	0	0	0	0	0	0	0	0	0	99	99	99	98	96	96	93	90	86
13	0	0	0	0	0	0	0	0	0	0	0	0	123	123	123	122	118	114	111	105
14	0	0	0	0	0	0	0	0	0	0	0	0	0	98	98	98	96	93	88	87
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	143	142	142	140	138	137
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	141	140	138	136	132
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	139	139	138	137
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	169	169	169
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	160	160
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	165

A.3.2 Reduced Data for Model II

The extraction from the Model I dataset yields the inputs R_T and F_T shown in Table 77 for Model II.

Table 77: Complete Model II data extracted from Dataset 2 for Model I.

Year	X_{new} [Components]	R_T [Components]	F_T [Components]
1	100	0	0
2	290	0	0
3	250	4	4
4	450	4	4
5	500	5	5
6	630	16	16
7	500	24	24
8	390	32	32
9	240	48	48
10	150	66	66
11	0	82	82
12	0	99	99
13	0	123	123
14	0	98	98
15	0	143	143
16	0	141	141
17	0	139	139
18	0	169	169
19	0	160	160
20	0	165	165

A.3.3 Dataset Properties

Like Dataset 1, the choice of Weibull parameters used to generate Dataset 2 imply two key characteristics:

- As with Dataset 1, the components are in the aging region of the Weibull bathtub curve (see Section 7.3 for additional details). However, the population

in Dataset 2 is aging faster than that of Dataset 1 since the shape parameter is larger than 1.5.

- The characteristic lifetime of the Dataset 2 population is only 20 years.

This combination of Weibull parameters addresses the other possible type of dataset that one could encounter for an aging population of components. In this case, all the components in one vintage fail by the end of the dataset as opposed to a maximum of about 50% of one vintage as in Dataset 1. It is important for both models to be capable of modeling these two types of component populations.

Figure 136 shows the total annual failures for Dataset 2. Note that the failures increase at a faster rate and to higher levels than was observed in Dataset 1. On the other hand, near the end of the dataset the total failures decrease slightly as the older vintages are replaced by nearer components.

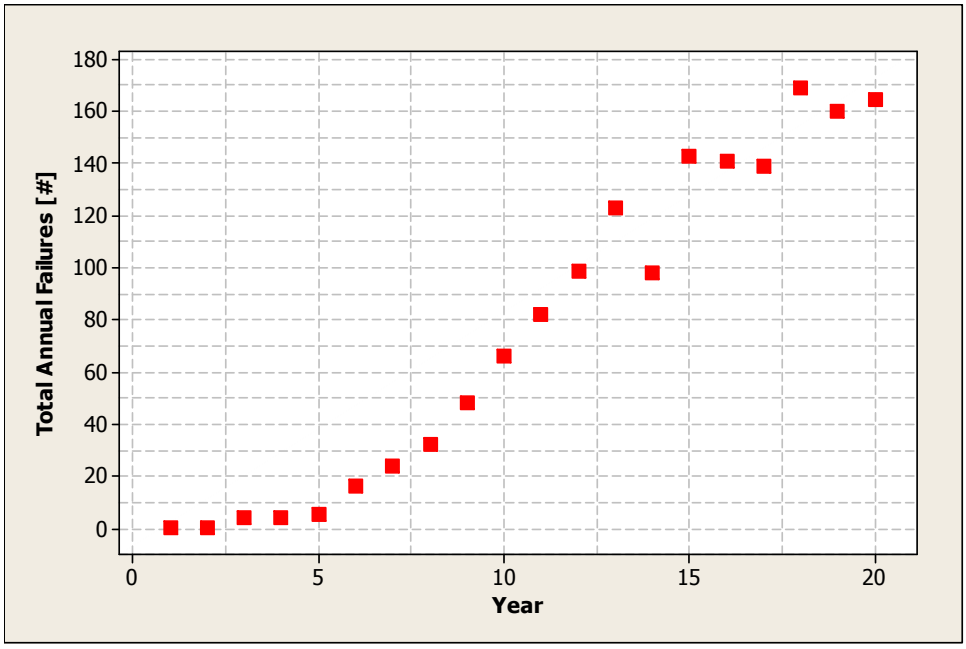


Figure 136: Annual failures for Dataset 2.

Table 78 shows the resulting annual failure rates.

Table 78: Annual failure rates for Dataset 2.

Year	In Service Population [Components]	Annual Failures [#]	Annual Failure Rate [#/100 Comp./Year]
1	100	0	0.00
2	390	0	0.00
3	640	4	0.63
4	1090	4	0.37
5	1590	5	0.31
6	2220	16	0.72
7	2720	24	0.88
8	3110	32	1.03
9	3350	48	1.43
10	3500	66	1.89
11	3500	82	2.34
12	3500	99	2.83
13	3500	123	3.51
14	3500	98	2.80
15	3500	143	4.09
16	3500	141	4.03
17	3500	139	3.97
18	3500	169	4.83
19	3500	160	4.57
20	3500	165	4.71

As expected, the failure rates of Dataset 2 are higher than those of Dataset 1. The maximum failure rate of Dataset 2 is more than 1.5 times the maximum failure rate of Dataset 1.

A.4 SUMMARY

This appendix has described the creation of two datasets that will allow both failure prediction models to be used on the same data. This will allow for comparison of the performance of each model as is demonstrated in Chapter 3 and Chapter 5.

REFERENCES

[1] BROWN, R.E.; WILLIS, H.L., "The Economics of Aging Infrastructure", *IEEE Power and Energy Magazine*, May/June 2006, pp. 52-58.

[2] BROWN, R., *Electric Power Distribution Reliability*, Marcel Dekker, New York, NY, 2002.

[3] BILLINTON, R.; ALLAN, R.; *Reliability Evaluation of Power Systems*, Second Edition, Plenum Press, New York, NY, 1996.

[4] NEETRAC, "Overview of Cable System Diagnostic Technologies and Application Overview for CDFI Participants", NEETRAC Proj. 04-211 report, Update 1, December 2006.

[5] DENSLEY, J., "Ageing Mechanisms and Diagnostics for Power Cables – An Overview," *IEEE Electrical Insulation Magazine*, Vol. 17, No. 1, January/February 2001, pp.14-22.

[6] QUAK, B.; GULSKI, E.; WESTER, P.; SMIT, J.J., "Fundamental Aspects of Information Processing and the Decision Process to Support Asset Management", Proc. of the 7th International Conf. on Properties and Applications of Dielectric Materials, June 1-5, 2003, Nagoya, pp.982-985.

[7] BROWN, R.E.; HUMPHREY, B.G., "Asset Management for Transmission and Distribution", *IEEE Power and Energy Magazine*, May/June 2005, pp. 39-45.

[8] OSTERGAARD, J.; JENSEN, A.N., "Can we Delay the Replacement of this Component? – An Asset Management Approach to the Question", Electricity Distribution, 2001. Part 1: Contributions. CIRED. 16th International Conference and Exhibition on (IEE Conf. Publication No. 482), June 18-21, 2001, Vol. 5, pp. 5.

[9] KOSTIC, T., "Asset Management in Electrical Utilities: How Many Facets it Actually has", IEEE PES General Meeting, July 13-17, 2003, pp. 275-281.

[10] BROWN, R.E.; SPARE, J.H., "Asset Management, Risk, and Distribution System Planning", IEEE PES Power Systems Conference and Exposition (PSCE), October 10-14, 2004, Vol. 3, pp. 1681-1686.

[11] LEEMIS, L.M., *Reliability – Probabilistic Models and Statistical Methods*,

Prentice Hall, Inc., Upper Saddle River, NJ, 1995.

[12] LI, W., "Evaluating Mean Life of Power System Equipment with Limited End-of-Life Failure Data", IEEE Trans. on Power Systems, Vol.19, No. 1, February 2004, pp. 236-242.

[13] QUIGLEY, J; WALLS, L., "Confidence Intervals for Reliability-Growth Models with Small Sample-Sizes", *IEEE Trans. on Reliability*, Vol. 52, No. 2, June 2003, pp. 257-262.

[14] International Electrotechnical Commission, *Reliability Growth – Statistical Test and Estimation Methods*, IEC 61164, Second Edition 2004-03.

[15] KIM, C.J.; LEE, S.-J.; KANG, S.-H., "Evaluation of Feeder Monitoring Parameters for Incipient Fault Detection Using Laplace Trend Statistic", *IEEE Trans. on Industry Applications*, Vol. 40, No. 6, November/December 2004, pp. 1718-1724.

[16] LU, H.; KOLARIK, W.J.; LU, S.S., "Real-Time Performance of Reliability Prediction", *IEEE Trans. on Reliability*, Vol. 50, No. 4, December 2001, pp. 353-357.

[17] ENGEL, S.J.; GILMARTIN, B.J.; BONGORT, K.; HESS, A., "Prognostics, The Real Issues Involved with Predicting Life Remaining", Proc. of IEEE Aerospace Conference, March 18-25, 2000, Vol. 6, pp. 457-469.

[18] GULACCHENSKI, E.M.; BESUNER, P.M., "Transformer Failure Prediction Using Bayesian Analysis", *IEEE Trans. on Power Systems*, Vol. 5, No. 4, November 1990, pp. 1355-1363.

[19] ENDRENYI, J., "Introduction to Maintenance", IEEE Tutorial on Asset Management – Maintenance and Replacement Strategies presented at *IEEE PES General Meeting,* Tampa, FL, June 24-28, 2007.

[20] ENDRENYI, J., *Reliability Modeling in Electric Power Systems*, Wiley & Sons, New York, NY, 1990.

[21] CHINNAM, R.B., "On-line Reliability Estimation of Individual Components, Using Degradation Signals", *IEEE Trans. on Reliability*, Vol. 48, No. 4, December 1999, pp. 403-412.

[22] BERTLING, L., "RCM and its Extension into a Quantitative Approach RCAM", IEEE Tutorial on Asset Management – Maintenance and Replacement Strategies presented at *IEEE PES General Meeting,* Tampa, FL, June 24-28, 2007.

[23] HERNÁNDEZ-MEJÍA, J.C., HARLEY, R., HAMPTON, N., and HARTLEIN,

R., “Characterization of Ageing for MV Power Cables Using Low Frequency Tan δ Diagnostic Measurements,” under review to be published in the IEEE Transactions on Dielectrics and Electrical Insulation.

[24] HERNÁNDEZ-MEJÍA, J.C., PERKEL, J., HARLEY, R., BEGOVIC, M., HAMPTON, N., and HARTLEIN, R., “Determining Routes for the Analysis of Partial Discharge Signals Derived from the Field,” Accepted for publication in the IEEE Transactions on Dielectrics and Electrical Insulation.

[25] BROWN, R.E.; MARSHALL, M., “Budget Constrained Planning to Optimize Power System Reliability”, *IEEE Trans. on Power Systems*, Vol. 15, No. 2, May 2000, pp. 887-892.

[26] PAINTON, L.; CAMPBELL, J., “Genetic Algorithms in Optimization of System Reliability”, *IEEE Trans. on Reliability*, Vol. 44, No. 2, June 1995, pp. 172-177.

[27] *IEEE Guide for Electric Power Reliability Indices*, IEEE Std. 1366™, May 2004.

[28] FORREST, W., “Predicting Medium Voltage Underground Power Cable Failures and Replacement Costs,” presented at the Western Electric Power Institute Underground Distribution Workshop, Portland, OR, April 8, 1997.

[29] KENDALL, M.G., *Rank Correlation Methods*, C. Griffin, London, 3rd Edition, 1962.

[30] ABERNETHY, R.B., *The New Weibull Handbook*, Robert B. Abernethy, North Palm Beach, FL, 1993.

[31] DODSON, B., *Weibull Analysis*, ASQ Quality Press, Milwaukee, WI, 1994.

[32] KAPLAN, E.L.; MEIER, P., “Nonparametric estimation from incomplete observations”, J. *American Statistical Assoc.*, Vol. 53, No. 282, June 1958, pp. 457-481.

[33] BILLINTON, R.; LI, W., *Reliability Assessment of Electric Power Systems Using Monte Carlo Methods*. New York: Plenum, 1994.

[34] LIEBER, D.; NEMIROVSKI, A.; RUBINSTEIN, R.Y., “A Fast Monte Carlo Method for Evaluating Reliability Indexes”, *IEEE Trans. on Reliability*, Vol. 48, No. 3, September 1999, pp. 256-261.

[35] BACA, A., “Examples of Monte Carlo Methods in Reliability Estimation Based on Reduction of Prior Information”, *IEEE Trans. on Reliability*, Vol. 42, No. 4, December 1993, pp. 645-649. MC

[36] BOX, G.E.P.; HUNTER, W.G.; HUNTER, J.S., *Statistics for Experimenters*, John Wiley & Sons, Inc., New York, NY, USA, 1978.

[37] PAPOULIS, A..; PILLAI, S. U., *Probability, Random Variables and Stochastic Processes*, Fourth Edition, McGraw-Hill, New York, NY, 2002.

[38] *IEEE Guide for Testing Shielded Power Cable Systems Using Very Low Frequency (VLF)*, IEEE Std. 400.2™, March 2005.

[39] LI, W., "Risk Based Asset Management – Applications at Transmission Companies", IEEE Tutorial on Asset Management – Maintenance and Replacement Strategies presented at *IEEE PES General Meeting,* Tampa, FL, June 24-28, 2007.

[40] PUTERMAN, M.L., *Markov Decision Processes*, Wiley, New York, NY, 1994.

[41] GOLDSMAN, D.; KIM, S.-H.; NELSON, B.L., "Statistical Selection of the Best System", Proc. of the 2005 Winter Simulation Conference. Orlando, FL, Dec. 4-7, 2005, pp. 178-187.

[42] BERTSEKAS, D.P., *Dynamic Programming: Deterministic and Stochastic Models*, Prentice Hall, Englewood Cliffs, NJ, 1987.

[43] KAY, STEVEN M., *Fundamentals of Statistical Signal Processing, Estimation Theory*, Prentice Hall, Upper Saddle River, NJ, 1993.

[44] FLEMING, W.H.; MCENEANEY, W.H.; YIN, G.G.; ZHANG, Q., *Stochastic Analysis, Control, Optimization, and Applications*, Birkhauser, Boston, MA, 1999.

VITA

Mr. Joshua Perkel was born in Portland, Oregon, on November 2, 1978. As an undergraduate, he attended the University of California, Berkeley, where he earned a B.S. in Electrical Engineering and Computer Science (EECS) in 2001. Josh began his career at the Georgia Institute of Technology in 2003 earning an M.S. in Electrical and Computer Engineering in 2005. Following the completion of his PhD, Josh will remain in Atlanta with his family and plans to begin his professional career with the National Electrical Energy Testing, Research, and Applications Center (NEETRAC). He will continue his work in asset management strategies for power system infrastructure.